

## Brainstorming session 1 Group 1

### Transcript

00:00:00 Speaker 1

My group here to jump in. We had a really wide-ranging and fast discussion. Under styles and genres of small data sets; one of the first tricky things that came up with was people saying: is style and genre really the kind of categorization that we care about, but also: what makes a small data set? And so there's big, kind of ill-defined questions there.

Does anybody want to say anything else about that?

00:00:26 Speaker 1

OK.

00:00:31 Speaker 1

I think the span of music that came up that we have either experience with or personal interest in, ranges from classical, Turkish makam, field recordings, oral histories, pop rock, indie sub-genres; music, you might talk about music that a specific person listens to; music that might be popular in a particular location at a time in history, right? So that as an example of something that crosscuts genre over style.

00:01:01 Speaker 1

For, let's see, features or characteristics of music, we talked about, you know, how is it represented: symbolic scores, audio recordings, the fact that some of these, you know, this music data in scare quotes, might actually be richly multimodal - it might include video of dance or performance; it might include speech; it might include descriptions of how data was collected and cultural context that it comes from, cultural identities of people involved, how it was collected, and that leads to one of the challenges, which is how, you know, how that gets represented or completely annihilated when you put that music into training data sets and the ethical implications of that, the cultural implications of that.

00:01:49 Speaker 1

Other challenges that came up are things, you know, both that I think are well known bias of, you know, large data sets towards Western music traditions; very practical challenges around things like OMR or dealing with audio datasets that have, you know, they're not segmented, there's lots of silence, you don't know exactly what's in it; copyright issues, accounting for the relationships between what we might call a small data set and the rest

of the musical context, or even all music. Somebody wrote, I really like this, sometimes representation is restrictive. Arguably representation is always restrictive. Business considerations, and we didn't have a whole lot around how data sets are collected and shared, especially like looking forward, people made a great practical list of people involved in these processes, representations used to these.

00:02:45 Speaker 1

What am I missing guys?

00:02:47 Speaker 2

Can I have a quick word on the electing?

00:02:49 Speaker 2

You can really come to that, but I just wanted to kind of mention the differentiation which is quite prominent in science between the so-called helicopter research versus the communities collecting their data themselves.

00:03:07 Speaker 1

I think it's a great point and actually that resonates with another project that I'm in, in the visual arts space where there have been really strong and I think well founded pushbacks amongst members of the community to calling the things that they're working with data.

00:03:25 Speaker 3

Thank you very much.