DATA AFFAIRS & Artifical Intelligence – Audiotranscript ENG

Prof. Dr. Birgitt Röttger-Rössler (in the following abbrev. w/ BRR) is social & cultural anthropologist and from 2015-2022 has been head speaker of the CRC1171. Currently she is head of the associated project "Das neue vietnamesische Berlin".

Prof. Dr. Rainer Mühlhoff (in the following abbrev. w/ RM) is associated professor for philosophie at Universität Osnabrück and researches on ethics, social philosophy, philosophy of technology and media science of the digitalized society. In 2023 his book ‚Die Macht der Daten. Warum Künstliche Intelligenz eine Frage der Ethik ist' has been released w/ V&R unipress publishing house.

BRR: I'm speaking with Rainer Mühlhoff, Professor of Ethics of Artificial Intelligence at the Institute of Cognitive Science at the University of Osnabrück. Rainer, I'm very pleased that you're sitting here with me today and have taken the time for our conversation.

RM: Yes, thank you for the invitation. I'm also pleased.

BRR: We want to talk about AI. AI has been in the public discussion for several months now, especially with the media hype around ChatGPT. However, AI has been omnipresent in our society for a long time, influencing our lives and fueled by our behavior in dealing with digital media. I would like to talk with you today about this omnipresence of AI, and in your writings, you speak of AI as a socio-technical system. Could you elaborate on this as a starting point?

RM: Yes, gladly. To perceive AI as socio-technical systems primarily means looking at the societal embedding of AI technology that already exists today and has implications for all of us. There's a certain definition associated with it, which is important to establish initially. In the public hype surrounding AI in cultural discussions and elsewhere, AI is often portrayed as a vision of the future, either utopian or dystopian. The dystopian view suggests that eventually AI will take over and subjugate us all. The utopian view suggests that somehow AI will lead to a world where we no longer have to work but are constantly cared for, and so forth. These are always future discourses. So, the first distinction is: No! AI technology is already here; it has implications for the majority of people on this Earth. When discussing AI ethically and politically, we should focus primarily on the current implications and not just the future ones. The second distinction is: In mainstream discourse, AI is often depicted as embodied entities, such as robots or self-driving cars. Anyway, it's always about materially opposing entities that talk to us or interact with us and shake hands with us and so forth. Then artificial intelligence is located within these objects or machines or entities. The point is: Most AI today doesn't confront us materially. It's information processing in information networks, in data centers. Most AI today deals with evaluating the data we constantly and daily produce when using digital media and services, learning from these data, and using them to treat us in an individualized way. Making suggestions, categorizing us into risk

groups, suggesting a route, and so forth. These are artificial intelligence services that aren't provided by an entity confronting us and interacting with us worldly but rather occur in data networks that are immaterial. Therefore, they are not visible and tangible. And the third point, which brings us to the socio-technical character at the latest... The third point is that this dominant and relevant AI technology today, a) in terms of its implications but also: b) for example, in terms of where the most money is involved. That's all technology that exploits our data. The data that we've been accumulating since the 2000s, since smartphones and internet connections became widespread, at large companies. It's no coincidence that we're experiencing this hype around 'machine learning-based AI' right now. Machine learning is precisely that, machine learning. It learns from our data. And these technologies... the ideas for them existed in the 20th century, but back then, there weren't as many data on which one could train good 'machine learning models,' models that can actually do something. So, the reason we're currently experiencing this hype around 'machine learning-based AI technology' is because we have the data. And why do we have the data? Because with the turn of the millennium, networked media technology became acceptable. So, the internet was released in the nineties, and smartphones were available from the 2000s. And these are the two major technologies that have led to data from all areas of life, from almost all people on Earth, being constantly aggregated. Only through these massively aggregated data sets have good and powerful 'machine learning systems' become possible. So, we all play a facilitating role, so to speak. And every day, by continuing to use our devices and produce data, we contribute to these systems. And therefore, they need to be analyzed as socially and culturally embedded.

BRR: You emphasize in your writings, especially in the text on the power of data, that a power-analytical perspective on AI is necessary and that the particular challenges for an ethics of artificial intelligence lie precisely here. Could you elaborate on that a bit more?

RM: Yes, I think what an ethics of AI can look like is quite exciting. It's quite contested at the moment because it's a young field. There are now professorships, but there's no canon. There's also no hegemony yet. I would say there's a lot of AI ethics being pursued in the paradigm of applied ethics. That means you have specific application domains and consider considerations there. So, I don't know, how should the care robot be designed to treat people with dignity or something like that. These are very important ethical questions that are dealt with in a domain-specific manner. What I'm doing is more of an ethics that is actually, honestly, a critical theory. For me, I'm even of the opinion that you cannot conduct AI ethics without a concept of power. Or from a power perspective, one must conduct AI ethics because the most significant structural, that is, societal-wide and society-restructuring effects of AI are indeed power effects. That is, I'm a bit more interested in things like: What structurally does AI do to our society? So what forms of exploitation, discrimination, perhaps even favoritism of some groups, disadvantage of other groups. Also, such global connections. That means global exploitation mechanisms. What kind of relation to the global south is inscribed in the technology and so on? That would be the perspective I take to then conduct ethics of AI. This is an ethics that is very close to political questions and wants to do very little of what could be termed checklist ethics. This is also a negative term.

So, a lot of AI ethics is in the form of, "Yes, we just want to have a checklist now. So that we can somehow make our AI ethically, we would like to be able to tick a few boxes. If all the boxes are checked, then we can sell the product!" I don't do that, and I don't find that useful either because I believe good ethics means that people also have to think for themselves and take responsibility themselves and not just fulfill, in the mode of fulfillment of external imperatives that press upon them. So, ethics, for me, is also a philosophical discipline that has a lot to do with taking responsibility, character building, actually virtues. And not just with something that, in extreme cases, even becomes 'white washing,' right? Ethics is often accused of 'ethics washing.' So: There is now a great demand for AI ethics because the industry actually says it's better for them than if AI were regulated. So, because ethics remains non-binding. There are hundreds of industry-sponsored 'ethics white papers' for various AI products because the industry also wants to produce the image that it's trying to make its AI technology and products ethical. And that's... it's analyzed as a major discourse strategy to prevent regulation, hard regulation, which also comes with boundary-setting. And I believe that when you conduct AI ethics or especially also teach it at university to students who are not just philosophy students but also computer science or cognitive science students, it's definitely my role to say that this is not ethics. But ethics would rather mean critically penetrating what AI does to our society.

BRR: Yes, thank you! That's very exciting and interesting, and I... Yes, above all, also to say: "Yes, ethics can be used as such a, yes, as a little tablet, as a protective sheet to actually operate relatively uninhibited behind it." Where are the power holders? Where are they seated? Maybe regarding that...

RM: Yes, exactly, thank you for coming back to that because that was the actual question earlier. I would say... I think it was necessary to address this briefly beforehand because now one can say... So, AI must be examined as a power phenomenon. But the power of AI is complicated. It's not simply accompanied by certain instances or people having this power. But I would say you have to look at at least two levels. And one level is a participatory level. That means AI as socio-technical systems, as collaborative intelligence networks, where all of us, by producing data daily on our end devices, contribute and are thus implicit parts of these power apparatuses. So, that would be the first level. And the other level is then a level of accumulation. So, of course, AI is accompanied by a form of power accumulation among large economic players. But I think it's very important to analyze these two levels in order to understand each other. So, let's start briefly with this participatory level. I think... it's really crucial that... there would be no AI without the millions of users who provide data daily. And by producing data, they also provide their cognitive performance, so to speak. So somehow... My favorite example is: Facebook, when it was still called Facebook, as the company itself was still called Facebook. In the 2010s, Facebook developed a facial recognition AI, 'DeepFace.' One of the first in the world. What do you need if you want to build facial recognition with machine learning? You have to train these systems. They learn from data, as mentioned before. And that means you need millions of labeled facial images. So images where there's a face, and the information of who that is. And getting such data, especially if there are a lot of them, is very difficult, or in doubt just expensive for such

companies. So which companies are most likely to do something like that? Companies that can persuade users to provide such data sets for them for free. And that's exactly what Facebook did by introducing this function in the early 2010s where you could label an uploaded photo. So, you could mark people's faces, right? When you upload a photo. Nowadays, all social networks do that. And this function was introduced to 'nudge' (to encourage) people to provide labeled facial images. That's, of course, the internal rationale of the company. The whole thing has to be turned into a social product somehow. And the social product was more like labeling your photo album, getting notified when someone else uploads a photo of you, being able to easily share the information that you were at this or that party on your timeline. So, it was dressed up as a kind of social interaction. So uploading a photo and tagging yourself on it is a way we can socially interact nowadays. And social media has contributed to making that one of the standard ways of social interaction, which you immediately notice when you observe teenagers, right? And the whole thing is a 'nice hack,' you could say. Somehow, a structure that captures data from all of us has been hacked into social reality. Social reality was, of course, also transformed by it - that's where we are at structural effects. And the crucial thing now is: We all make that possible. We can't say that we are purely passive parts or purely passive actors in this game. I wouldn't say that individual users should be held responsible or blamed individually because they labeled faces, so Facebook could build a facial recognition AI. That would be a bit too much to say. So, a structural problem also requires a structural solution. But nevertheless, a structural constellation means the interaction of many millions of actors, which somehow enables a parallelized interaction. In this case, of this AI system. And by labeling a face, we contribute a small portion of cognitive performance. So, recognizing faces has always been a very difficult problem for algorithms. So, an algorithmically difficult problem - very easy for humans. And the solution to this difficult problem by AI lies in the fact that we have created an information network, a global one, through social media and our phones, where we can access, portion-wise, the cognitive resources, cognitive performances of people, users, which are all fed into or orchestrated by this AI system, so that the AI system can then recognize these faces.

BRR: Mhm, thank you! I think this example really clarifies what you mean by socio-technical systems, and wonderfully demonstrates the intertwining. So, the impact on society, on traditional communication modes that are now altered by the influence of these possibilities and so on. Um (...) What I would like to address again in our conversation today is that I would like to incorporate, integrate our conversation into the e-learning portal 'Data Affairs'. And while we were actually creating this portal, focusing on data management in ethnographic research, we kept thinking and discussing as a team, saying: "Somehow we need (...) We need to tackle this issue!" Even if it's not immediately related to concrete data management in social science research, it's somewhere behind it. And what we want to do is to sensitize, with the portal, aspiring researchers, anyone interested, to think about what they do with their information, with their so-called data. Not just shooting them into clouds, but considering privacy, data security seriously. And (...) thinking about it involves ethical questions as well. So: Does everything necessarily have to be stored in digital repositories? Is that necessary? That's one thing. And another thing that always crosses my mind, um (...)

Yes, this whole data management discussion is based on these so-called FAIR principles. So this wonderful slogan, which is actually somewhat ambivalent, right? So it actually means: Findable, Accessible, Interoperable, and Reusable. So it makes knowledge, what you have acquired in very complex contexts, into small, processable, discoverable, reusable units, right? So it breaks it down somehow. Yes (...) And wouldn't a critical look also be meaningful? And (...) Some in these contexts also speak of 'datafication', or 'data extraction', or 'data mining' and so on. So, there are indeed very critical voices in the scientific community about this as well. Yes, that's just a ball I wanted to throw back to you. What would you say to that? Or your opinion on this topic? On this discomfort? Is it actually a discomfort that I, that our team has? RM: Yes, I can relate to that very well. And from my research perspective, I can also clearly say that I see very great dangers there. And what FAIR means can be the starting point for enormous unfairness, so for social consequences and effects that have a lot to do with unfairness or discrimination. I believe one should be aware that today (...) So we live in an age where mainly mass data, anonymized mass data, are interesting. That means promising people or the people depicted in this research data their anonymization. That does not prevent the powerful use of these data sets. It's not about, or it's not the greatest danger nowadays, to re-identify an individual in these data sets, to breach the promise of anonymization. Which is often possible and which is also a big problem in itself. But even if we now assume perfect anonymization and the individuals are not re-identifiable. Then we still have valuable resources with the data sets, from which especially machine learning systems can learn to distinguish people of different kinds. And that's what we're interested in. We're interested in being able to automatically classify people into different categories or social boxes, in order to treat them differently. So when they make research data publicly available, collected in the humanities and social sciences, one must assume that the insurance industry can and wants to access this data. And they definitely want to. Or, let's say, a company that builds AI systems to support job selection processes. And such systems are particularly interested in subcultures, minorities, and their social lifestyles, because they want to be able to automatically recognize whether someone belongs to such a minority or somehow risky lifestyle or with a presumed risk associated with lifestyles. That means, opening up such data, and as I said, we're talking about anonymized data here, potentially exposes vulnerable groups to an even greater risk of discrimination. So especially, especially when dealing with research that studies such groups. Even if you're dealing with research that focuses more on, I don't know, whether such a thing exists, the majority society or people who think they have nothing to hide, or who don't belong to any minority, presumably (...) Even this research is risky for society as a whole. So these data are risky for society as a whole, because you also need the data of the many supposedly normal people to be able to recognize the 'a-normal', in quotation marks, in contrast. That means, all the people who supposedly don't have an individual risk by disclosing data also contribute to the fact that other people can be discriminated against.

BRR: Yes, these (...) So, as you said, important for normalization!

RM: Exactly! They set the standard.

BRR: Yes, yes. So standards are ultimately always quantitative calculations. What the majority does is the norm. Or what the majority can, or health-wise also, right? So, these are always majority perspectives that set the norm. RM: And this trend is totally supported by that. And I want to, I want to add one more thing! A) We completely lack awareness that there is significant societal risk in the reuse, in the secondary use of research data. So, when we collect data - these are often experiments on humans or in the field - ethics committees are always involved. And these ethics committees assess the primary purpose of this data collection. So: What do you want to research? Is it ethically justifiable? However, they never assess the possible secondary uses of this data. And if the trend now says that research data must be made available in publicly accessible repositories, then that means that we must anticipate the entire spectrum of imaginable and even unimaginable secondary uses, especially by completely different actors. By actors that are not calculated with, actors that definitely have more discriminatory or abusive or exploitative intentions. You would actually have to take all of this into account when considering whether to collect such data. That means, the limited view of the primary purpose would actually be circumvented. And that would be a fundamental step. So, it would actually fundamentally overhaul our entire ethical evaluation system of such research. And my opinion is that the researcher who is in the field actually assumes a new form of responsibility there. So if you are in anthropological or ethnological research gaining access to a field. These are often spaces where it's not necessarily obvious that someone can stick their nose in there. Then responsibility comes with it. You were brought into confidence and held responsible by the field or by the people you are allowed to investigate. And (...) this responsibility is inseparable from the researcher standing in the field. And the moment she hands over this data to a repository and exposes it to uncontrolled secondary use, she violates, in my opinion, this responsibility, this trust placed in her.

BRR: Exactly, that's an argument that often comes from social and cultural anthropologists. There is generally such a discomfort. So, we also conducted focus group discussions with colleagues, and there is this discomfort exactly with the criteria and aspects that you just formulated. Now corresponding repositories, which are also available for social science, qualitative data and so on, say: "Yes! We secure the access! It's only available upon request!" The context is always very, very important, they say, to ensure that this data material is not isolated, can be viewed in completely different contexts. But I think doubts are in order. Who can guarantee me that? And can these repositories guarantee that with their limited access, with all their control systems? Can they guarantee that? And: The stuff is on the net and then accessible somewhere, and can be used completely detached from the collection contexts. And (...) So for me (...) Yes, I just have these question marks, and what you just outlined supports that.

RM: Exactly! And we not only have to think about hackers or people who gain unauthorized access to the data on the net, but there is no legal regulation that restricts access to such data, for example, to only certain research. So, for example, data protection does nothing if it's all anonymized data. That means, you can only rely on the promise of the operator of this repository that they will not only today, but also in ten or 20 years, apply appropriate criteria

in deciding who gets access. And you can (...) There is no legal framework for that, only this promise. So this bilateral relationship between the operator of the repository and you, or the subjects in the data. One has (...) there's this example, it's just about four weeks old. (...) In the UK, this 'BioBank project' was launched in the 2000s. About 500,000 people voluntarily go to the doctor at relatively regular intervals to be examined for certain health characteristics, so that a time series, a very long time series of health data from volunteers is created. And these volunteers (...) So it's about cancer, but also about behavioral data, substance abuse, lifestyle, and so on. And these people were told in the 2000s, when the 'BioBank project' was founded, that this data would only be used for medical research and would not be passed on to, for example, the insurance industry. And it has now turned out: This data has been passed on to the insurance industry. Anonymized. That's why the people in the datasets also have no claim. Data protection is not violated if it's anonymized. The point is: The insurance industry is not interested in personal data, but in large datasets of anonymized data. Because what they want, and what they have shown they have done is, for example, to examine the correlation between certain lifestyle patterns and, for example, the risk of having cancer. That means, what the insurance industry wants is, when you apply for, I don't know, a new insurance, then they look at it like this: What is your lifestyle? And they want to make predictions based on that. How risky are you? Or do you probably have certain diseases, in order to offer you the expensive insurance in such a case? And these datasets were used for this purpose because, namely, the regime, so the board of this Biobank - which is a company - this board, they just changed and the criteria for accessing this data have just changed over the past 15 years. And that's exactly what can happen with any research repository that is not subject to government regulation. And by the way, even if they are subject to government regulation, that can change too. We also have to always think about scenarios of power takeover by authoritarian or racist policies or political regimes, right? We think: Today it's not so risky when we collect data about minorities and store them in repositories for all time. But we can, I think, well imagine political developments in the future with these data doing things that we definitely do not want today and hopefully not in the future either. So we also have to think about that.

BRR: Yes, I think these are all very, very important aspects that we have addressed here and hopefully will make some people think. I would like to come back to the aspect of sustainability at the end of our conversation. These huge data centers that exist, which are growing and growing and getting bigger and bigger, in order to handle these data volumes, they have an enormous energy demand. And I think too little is being said about that in my opinion. Um, I don't even know if you can say something about that!? You probably have an opinion on that...

RM: Yes, that is currently an important debate in AI and especially in AI ethics research. That (...) AI technology is incredibly resource-intensive. Computing such models costs a lot of computing power, and computing power is always associated with energy expenditure. So power plants have to be built for that. When you think about what it costs, for example, to train such large language models, then it's several million euros. Most of which is the energy requirement, right? So essentially that's the electricity bill. So we're dealing with an

incredible scale here. You have to be aware of that. By the way, not just energy, but also rare earths, which are always associated with corresponding mining and economic exploitation relations, often with countries of the global south. So the extraction of corresponding materials. Extracting corresponding raw materials is inherent in AI technology, or digital technology in general, including the technology we're using to record this conversation right now, and so on. My role, when this question comes up, is - after I've said all that - to always point out that indeed, sustainability also has a social dimension. And that is often forgotten within, even within sustainability (...) So, sustainability is often forgotten, but within sustainability, it's often forgotten that sustainability also has a social dimension. That means a social component. Among the 17 UN Sustainable Development Goals, there are seven that relate to social goals. Things like equality, freedom from discrimination, access to a job, and so on. And AI technology is the technology that is not sustainable or must not be sustainable, especially from a social perspective. And we just talked about that. We just talked about whether research data, for example, has anything to do with the study of social lifestyles of minorities or certain societal groups, whether we should store them in publicly accessible or more or less accessible repositories for a long time. And we don't know what will be done with this data in the future, and it's very likely that this data can be used for discriminatory purposes against these people (...) So against these societal groups. And that's a point that directly violates the social sustainability goals relatively simply and directly. And this dimension must also be included, especially when it comes to research data. Research data is always about the debate: We want to somehow keep and open up these data for yet undefined, uncertain ideas for secondary use. And thus, a future aspect is addressed here that could go very wrong in terms of sustainability. And not only because of the energy consumption of what is then calculated, but also because of what is then done with it, so the purposes associated with data use can be socially harmful, especially if the data fall into the wrong hands or if the ethical evaluation or political regimes change, which decide whether the data may be used or not. The BMBF (Federal Ministry of Education and Research) is currently working on a legislative initiative for a Germany-wide research data law. It's in an early stage of conception. There is a key paper, and there are ideas in it like a right to use. So, the idea is currently being discussed that industry, private actors, should have a right to use, access to research data from public research institutions. That's in such a key paper. From my point of view, that's serious. Not only because of the asymmetry. Namely, the research data of private companies should not be opened up in this way, there is no access right provided for in the current discussion. But the research data of public institutions. And public research institutions are exactly the places where research, especially about and for societal minorities, must and should be done. Precisely for that reason, I would say that this data is particularly vulnerable and deserves special protection against exploitation by economic interests or directly discriminatory interests. Therefore, I think it makes a lot of sense to bring in the social dimension of sustainability and the 'UN social sustainable Development Goals' to argue for not storing all research data in repositories for all eternity.

BRR: Exactly, and that, yes (...) that it can actually be an aspect of sustainability not to do so. Rainer, thank you very much for this conversation!

RM: Birgitt, it was a pleasure! Thank you too!