A1  
 That sounds good. Yes. Exactly, and then we can start the interview. We just got a short introduction regarding your work. To be a bit more specific now, how is machine learning or AI being used there?

I11  
 Well, a lot in the direction of classification stuff. Right now, this whole LLM topic is getting quite big because many want chatbots, many want to enrich their documentation with large language models so you can basically talk to the documentation. That is a big topic at the moment. And yes, what really counts as machine learning basics is mostly classification stuff, clustering stuff, and anomaly detection. Those are often the topics now.

A2  
 I… could you briefly explain again, for me with no technical background, what is XXXXX X? What is the product or what exactly do you or your team do?

I11  
 We are basically a service provider. Not basically, we are a service provider. For… I always call it digitalization, data science, software development. So topics that are in this whole data and digitalization environment. We basically go to companies or companies ask us if they want to do something with data, very plainly. It could really be that they come and say, we heard you need AI now, let's do something with AI. And they have nothing yet. But it could also be that they have been collecting data from their machines for years and now want to see what can be done with this data. It could also be that they already have very advanced things and have a small idea they want someone outside to try out a bit. There are many constellations. But a classic case is that we have a machine manufacturer, for example a company that manufactures facilities that make films or paper. These are house-sized facilities. Facilities cost millions of euros, world market leaders, all of them are world market leaders with us. And now they want to equip these machines with sensors so that from all the machine data you can see whether the product coming out at the end, for example the film or the paper being produced, is good. Because these are processes. There is not someone on the 13th floor looking at the film saying yes, it is good or bad, it all has to run automatically. Connecting all the processes with data, that is one case. And then maybe in the end they want the machine operator to have an app that shows yes, the last 500 meters were good, something is bad here right now. It all has to be in one system. Customers should be able to view all their machines. So topics like that are part of what we do among other things.

A2  
 Thank you, so I know you introduced it briefly, but I had to hear it again a bit more precisely for my understanding, thanks.

A1  
 Exactly, what would interest me in this context is your target group of companies you want to reach. Is it limited or is it the entire sector, no matter who has a problem with digitalization and automated data processing, you take care of it, or is it already manufacturing, for example, the example you just gave would be industrial manufacturing of a specific product.

I11  
 That is the core, which historically grew that way. Because, well, we have a few really strict exclusion criteria. That is like defense-related or ethically questionable stuff. Anything involving personal data we do not do. So we are more interested in machines and processes. We do not do online marketing or target group focusing. That is not our area at all. We also have not done financial services. Simply because it was never the area with this physics background. We are not excluding it for ethical reasons, but it is just not our area. We have a few medical technology projects. Yes, sometimes other things come up from the side. But the focus at the moment is really the German market and the machine and facility engineering sector, mostly medium-sized companies.

A1  
 Alright, okay cool. What are your everyday work processes like? And, as you said earlier, you also got a bit into software development. How does AI play a role in your everyday work processes?

I11  
 Well, there are two sides to it. One is what I actually do every day, the technical side. The other is what customers want, what you have to sell. Sounds silly, but the general understanding of AI and what AI is supposed to do for the German economy is miles away from what you sit at the computer and really do. What I actually do, if you come from research, is really the basics. We do not train large new models. We use a lot of what is already available. For smaller tasks, we just use SKLearn or something. So we are mostly in the Python world, training a small random forest or doing a bit of clustering. Really basics. In the LLM area we have also dealt a lot with local models, how to run them, and how to work with them. But we are not building crazy neural networks ourselves or worrying about architectures or which loss function is best. That is not our focus, we are really users. Yes, users of what is currently available. And often in the cases we have, and for what is actually needed in the industry, that is miles away from what is done in science, at what level people operate, what is needed, what is important for the process. So technically, really basics. If you were writing a PhD in machine learning, you would probably smile tiredly at it. The other side is the understanding our customers have, what they want, and what you have to sell. If you mention the word AI, a whole spiral starts somehow. Oh, now XXXXXXX is gone

A2  
 Yeah, I don’t know either. But we are recording and just keep going.

I11  
 We’re recording, then I can keep talking. Right, so this idea of AI, where somehow what you deal with, it’s often the case that small things, like a small classification, and I have a dataset here and I want to classify whether my product, that I produce, was good or bad, roughly. And you can often do that, I don’t know, with a small Random Forest. Then for the clients that is already huge AI. Hello, there you are again. You’re still muted.

A1  
 Okay, sorry, yeah, no idea, there was apparently a short university dropout just now. Did the meeting continue?

A2  
 Yes, we continued very smoothly…

A1  
 Then we already had a good fail-check here that we hadn’t had so far. Yeah, sorry. Yeah.

I11  
 Exactly, I was just saying that for our clients, when you do something with AI and you train a small, small Random Forest, that is often already, for them, big magic and has a big impact. And yes, you sometimes have to tell a bit more, and for them this world is a bit more flowery, more magical, than what you are used to from research.

A2  
 So if I understood correctly, you or you all build your own systems but also somehow use other systems or build on other things integrated in your setup

I11  
 Exactly. So now specifically in this, let’s say, LLM topic, these are just models and neural networks, roughly speaking. They are so big and you need so much data to train them. You can’t just rebuild that, it’s somehow pre-trained. These are large systems, supported by a lot of hardware. You can basically plug into them and use them, not rebuild them yourself. And there we use ready-made things. It’s always a big topic with data protection and so on, and another issue. Exactly. And for smaller things, where you want to experiment a bit, there are also many things you can really build locally yourself, that you can basically make from scratch yourself, where you don’t need to plug into external systems.

A1  
 Especially because many companies don’t even have such amounts of data for it to make sense. By the way, I’m also involved in an industrial project where we try to somehow bring AI into industrial manufacturing. And yes, the data there is also not like what we, as machine learners, know: a perfectly curated dataset, maybe not perfectly curated but already very structured. There is a lot of Data and Business Understanding first, which you have to somehow do in order to get to the point where you can train anything at all.

I11  
 That is actually often 80 percent of the work we do. Especially in initial projects, where we get to know the client, get to know the processes. It’s never the case that you arrive and 80 gigabytes of perfectly prepared data are just there to use. There’s a lot of preliminary work. That is the main work. So when I get to a point where I can fill something out somewhere fit, I’ve already worked many, many hours for that, roughly.

A1  
 Yeah, definitely. But super cool that you basically do both, because I also just had, we already had it a bit in the first interviews, there are many people for whom AI and Machine Learning now only means LLMs. So there are whole companies doing AI, but they only do LLMs and chatbots, which is fair enough, because that is currently also somehow what is most...

I11  
 Needed everywhere, everyone is getting into it.

A1  
 And I think it’s also hardly possible as an AI, Data Science, Machine Learning startup team, whatever, to work without completely ignoring the topic and saying: We’re not doing anything in that direction. Because then you basically give up a lot of the market

I11  
 Definitely. So, you can’t really, that’s not what we mainly do. And we also have a bit of a personal principle not to sell AI as a magic black box that will solve all your problems, but we really take a hands-on approach, first trying to understand the processes and then using AI or machine learning, in our terminology, only when it makes sense. We don’t go and say we sell you an AI solution that definitely includes AI and then it ends up being a linear regression or something. Instead, we first look at where it actually makes sense. There are machine learning topics in there, that’s kind of our approach.

A1  
 Yeah, very cool. Cool that you do both and also go in-depth. What are the biggest challenges for you, where you see that providing AI solutions or ML solutions, using the terminology interchangeably here, is difficult? What are the main challenges in providing or developing such systems?

I11  
 So, if we leave out the human part, meaning communication with the customer, expectation management, that’s one thing. Technically, it’s definitely data quality, I would say, data quality and also existing labeling, because of course processes in the industry were not set up from the start in a way that every good production is labeled as good production and every failure or bad production is labeled as bad and why it is bad. You often have, when we talk about classification, we had a case where it was about contract termination and there was a whole dataset with 100,000 contracts, like when they were signed and with which parameters, without personal data, and then we were supposed to somehow see when a contract ends. And then we had exactly one event where a contract had ended. So situations like that, where there is basically no data basis to do anything, those are the kinds of issues.

A2  
 And what would interest me now is, when you or when you bring in systems that you didn’t develop yourselves, how does the decision-making process happen and what are the challenges you notice there?

I11  
 So the decision process is very much driven by what the customers want. Especially regarding the LLM topic, because that’s what people are talking about a lot right now, using external services. You have to see where they are already operating. Microsoft Azure is kind of the environment where almost everyone or very many are already active. Microsoft made a deal with OpenAI so that you can use the ChatGPT or GPT instance locally, basically in your own environment. I haven’t fully figured out technically how that works. But the thing is, customers often already trust Microsoft because they live in the whole Office world and then say, okay, then we can also use the GPT model there. So it is somewhat determined by them. Otherwise, yes, we see what works. We try to work open source wherever possible, in our fundamental spirit. Of course, with topics like LLMs and chatbots, you eventually hit the limits with small local open-source solutions. If it is something that should run productively for a customer, you have to make a decision. And yes, I can’t really say how it comes about, but it’s often in agreement with the customer, with what they already have, and wherever possible, open-source and free solutions that make sense.

A1  
 Okay. Can you think of specific things where you would say, that would be a huge enabler and would drastically reduce some of the challenges or everyday problems?

I11  
 For the development of machine-learning-driven applications, do you mean?

A1  
 Yes. Either development or deployment. I would divide it like this. Either you develop your own software to solve a specific dataset or problem for a customer, or it’s really the case that you bring in solutions from OpenAI or similar and provide them. Those are kind of the two paths you have at the moment, I think.

I11  
 Yes. So clearly, in the processes, so now just hypothetically, if you naturally had good data, that would be great. Clear case. But in that sense, it is often not feasible, simply because it didn’t develop that way historically. So if, in principle, you thought about these kinds of topics, basically, when you set up machines and start taking data, if you already thought ahead that you might need labeled data, that you might somehow structure it, that is, put it into a data infrastructure, where even an external person, who doesn’t fully live in the client’s IT world, could get something out of it. Those kinds of things are often just hurdles, but I don’t think there is any product that really wants to solve this. So it is again the person designing the product that wants to solve it, designing the 101st product that should solve it but then doesn’t. Yes, those kinds of things, and the rest is often just a lot of research, because you have to piece it together yourself, from the hardware to the model you then need, the one you want, if there was the ideal research tool that could piece that out for me, but the world is also so fast-moving. If I look at something in three months, just like with this LLM topic now, we really started about a year ago or so, downloading a model in the console and rebuilding it somehow, and had to go into the architecture, and now you somehow have three lines of Python code and can do everything like that. It is just super fast-moving.

A1  
 Yes, definitely.

I11  
 So yeah, I think the researching, piecing things together, and staying up to date, nobody can do that for you. So those are the points, I would say.

A1  
 Yes, makes total sense. In our study, we also have a concrete possibility, kind of in the sense where one could maybe help. It goes in the direction of model selection or reporting of machine learning properties. Because that is at least something that has repeatedly caught my attention in contact with companies, that this communication about machine learning is a huge problem. And the ML scientists want to see the source code or at least the paper and all the formulas. But that is again something you cannot show to someone who doesn’t have this domain expertise. And we are trying a bit to bridge the gap, to close the gap between developers and experts and non-experts and users. Or also companies that develop AI solutions or are trying to integrate AI into their solution but don’t yet have this extensive domain knowledge. I will briefly share my screen for that. And then we can talk a bit about what I want to show. I just need to start the presentation. If PowerPoint cooperates. Hello. Okay, very good. That was it. And now I am showing my corresponding screen. That would be this one. So, can you see my screen?

I11  
 Not yet.

A1  
 I can maybe

I11  
 Ah now. Even in full screen, amazing.

A1  
 Very good, that is what I hoped for. Perfect, exactly. This is now just directly the example we want to show. This would be an AI label, as we have kind of drafted it, I would say. And before I explain anything in detail, I would like to ask for a spontaneous reaction. What do you see, what comes to mind?

I11  
 I see, it reminds me of the Nutri-Score. I only realize at second glance that it is not actually a well-known label. So yes, okay. So it is supposed to be, I see it as a label if I have a machine learning model, which tells me how much, what do we see, what it costs in terms of power, how fast it is, how accurate it is, and how robust it is. Yes, that makes sense. Very nice.

A1  
 Very nice. That is good. That is a nice, we can put that in the paper. Yes, exactly, completely correct. We have some detailed properties at the bottom, concrete metrics, and an overall scoring, which is visually somewhat inspired by the Nutri-Score. It is actually not so easy to create a new label without people saying it looks like something else.

I11  
 Yes, I understand that. And keeping it simple and

A1  
 Yes, so the Nutri-Score is pretty much the simplest thing you can do with a five-level scale. But because of that, it also immediately looks like that. Yes, exactly. When you look at it like this, do you maybe already see some kind of bridge to your topics or problems? Does it have any practical relevance for you? Potentially?

I11  
 Well, maybe eventually, when we get to the point where we even want to optimize something based on that. In what we do daily now, if you take it all together, the step at the end is really choosing the right model and then maybe checking what costs a little less and is a little faster. But we often don’t get that far. We spend a lot of time just prototyping and first checking if anything works at all. So it’s definitely good if you actually get to the point where you want to optimize things on that level. Honestly, in our work, that doesn’t happen very often. It’s not like in every project we eventually get to the point where we think about the machine learning model and its performance.

A1  
 Instead, the bigger hurdle is actually getting something meaningful out of the data and delivering a proof of concept. Optimizing it afterward would be more of a downstream step. How is it with LLMs? I mean, you’re always facing the question, which OpenAI model should we use for the task?

I11  
 Yes, actually, from what I’ve experienced so far, it has been the case that people said, no way OpenAI, we can’t do anything with it yet because of data protection and so on. So I’ve mostly been focusing on how to handle it with local models. Of course, nailing down accuracy is a bit more difficult in terms of performance.

A1  
 Especially given the client situation. You might be able to do it for some benchmark, but whether it is similar with client data, you just don’t know.

I11  
 Yes. And if I’m not aiming to have a linguistically nice, grammatically correct sentence and don’t want prose text, but something else, that makes it a bit tricky for choosing which external large models. Honestly, we just use whatever is available. If something is available because they already have a GPT in their Azure instance, that gets used first. It’s fairly pragmatic. Because the models you can buy are so good that for what we do, it almost doesn’t matter what you have, because they are usually good enough for the very specific task we have.

A1  
 Yes, alright. We actually brought a second label, too, so the idea is a bit like instead of difficult-to-understand papers or implementation details, you can communicate practical properties in a simple, understandable way. Here’s a second label as an example. Can we maybe make some comparisons, or can you do that?

I11  
 I see you adjusted the colors nicely from red to green. So if I now see a greener accuracy, a greener icon, I immediately see it has better accuracy. And I assume at the top there’s some kind of average of the different properties. Yes, I would say that makes sense, and I can see at first glance. I would see at first glance, if I know I need this accuracy or I have, I don’t know, time issues because something has to run fast, then I would see what it does.

A1  
 Yes, exactly, that’s kind of the idea, to represent it abstractly. I mean, of course, if you have the results in a Pandas DataFrame or something, you can also just sort based on columns. But even tabular representations, where you sort by a criterion and say we want best accuracy, it’s always a trade. You’re from the domain too, I don’t need to tell you that more model accuracy usually trades off negatively with model size, and in the end, you always have to weigh what is the perfect solution for our problem, given also the project or company priorities and which properties are valued. How is that for your understanding? Are there things on the labels that are completely unclear to you or where you say, I don’t understand what this means or how it comes about?

I11  
 Uh, no, actually, yes of course, you could ask how you determined accuracy, what algorithm is behind it to get the numbers, sure. As for robustness, the question would be, what is 0% and what is 100%, otherwise it all makes sense.

A1  
 Yes, probably you’re already thinking too much in methodological depth and maybe even thinking, I would rather see the implementation, how did you calculate accuracy, that would give me more insight. The goal, of course, is also to build this bridge to people who don’t have that expertise. If you try to put yourself in that role, do you feel this is something that could also help in your communication with the client?

I11  
 Oh, definitely. They are totally into images and colors and if… well, with categories like red, yellow, green you can always catch people, in my experience. So definitely. Of course, most of our clients, we also sometimes deal with people who are very much on our level, with whom you can determine something like that or discuss it. For some, of course, they don’t even care about that, roughly speaking. They just want to know what comes out of it and what the added value in euros is. But sure, with people on the other side, who might be a bit into these data topics but not very deep, it definitely helps when you are at the point where you need to optimize something and think about what you are using.

A1  
 Yeah, that’s actually quite nice. For management, you basically only need a red dollar at the bottom for how expensive it is to run the model and a green dollar for how much money we can earn with it. That’s the trade they are interested in, understandably. Cost and benefits, like that. Yes, exactly. Maybe just as a quick background, the labels, not exactly one-to-one, but very similar labels, I wrote a software for generating them fully automatically from a Pandas DataFrame. That means, if you feel like you could use this in a meaningful way, you can also send me an email, maybe I can provide something like that on GitHub for you.

I11  
 I’ll do that.

A1  
 Just as a side note. Do you have any ideas on how the label could maybe be designed differently or better? We’re not doing a user experience study here, but there might still be some learnings that can be discussed a bit.

I11  
 Well, one could maybe go further, like showing these values or percentages in bars, so that you see the number differently than just in color coding. But I think that’s my perspective as someone who deals with this every day. That’s tricky. For people like me, it helps to see whether I have 60 or 40 percent if I have a bar or something from 0 to 100. That’s my way of working. But for people who can’t get much from it, the color coding helps a lot more. And therefore I really like the icon for robustness. I need a moment to think, but I wouldn’t know how to do it better otherwise. Otherwise, I think it’s very good. No, otherwise.

A1  
 Yes, so… yes, XXXXXXXXX.

A2  
 Do you feel that, if you had to work with it, say you had to make a decision based on this AI label about which system to choose, which information would be important to you that might be missing, or would you say this is what I need? Of course, it also depends on the task or the client, I see that, but maybe something comes to mind.

I11  
 Well, yes, time and accuracy are the things I am always concerned with because if something takes too long it’s annoying. The power draw is also very interesting if you scale the whole thing up. What’s missing, or what is missing in practice if I really sit in front of it, are things you mostly find out only by trying: how well it actually works with the problems, the data I have, how easily accessible it is, from which packages it comes, how well those packages are maintained, and so on. That all plays into it. If I go on GitHub and see projects that trained great models or wrote architecture four years ago and nobody touched it since, those things matter too. That’s hard to integrate into a label, I think, but they are additional information I consider.

A1  
 Yes, that’s actually an interesting point. I mean, we do show the date at the top of when the label was issued. As it is now, the model is loaded and measured once. So at least there’s a little bit of time information, how current it is, similar to the last commit. Of course, that doesn’t guarantee reproducibility one-to-one. It could already be deprecated. But I know from my own experience, I look at the repo and if it has a star but hasn’t been updated in five years, I can’t really use it. Compatibility is limited and technology evolves fast. That was also feedback I received in a paper review, that you definitely need a rough time reference on the label.

I11  
 Yes, that makes sense. That’s also something I pay close attention to.

A1  
 If you look at it this way, where information is concerned, it doesn’t seem like the point is necessarily to come up with the perfect solution to the problem in everyday life, but rather to really understand and model the problem. But if you were in a situation, or sometimes are in a situation, where you have to make these decisions, like what to use here best, where do you get your information? What resources do you use?

I11  
 First of all, I ask because the people at our company, they are all physicists or many of them, so that is naturally the first step. Then, if I have something, like recently I had a case where someone wanted a time series forecast, and the last time I worked on something like that was in 2018, so a few years had passed, and a lot had happened. We basically started from scratch again and Googled, looked at Hacker News a bit to see what was going on. But then you don’t have much time, you can’t research for three days on customer projects because they want the result tomorrow. So you have to rely a bit on your gut feeling, honestly. Then you see what initially makes sense, try it out, and if it still seems reasonable after the first trial, you go with it for now, wait until it breaks, see who complains, and then continue. That’s kind of the flow.

A1  
 Exactly, that makes total sense. We actually gathered a bit about what other information sources could be used. Of course, scientific publications, for example on arXiv, but also elsewhere. Right now with Hugging Face, LLMs, etc., it often happens that the implementation and the interface also have some kind of model card. That originally comes from Google. Then there are benchmarks and leaderboards, for example on Papers with Code, where you can rank how good a model is on a certain dataset. Of course, there are also blogs, Hacker News which you already mentioned, to get an overview, library documentation. IBM also has the concept of factsheets, although those are really used only for IBM-internal models. They don’t have public software. Here are examples for MobileNet V3, for which we saw the label earlier, for the underlying publication, a model card, the overview page on Papers with Code, blog posts on the topic, Keras documentation, and an IBM factsheet, even if it’s not the one for MobileNet. When you look at all these forms of documentation side by side, where do you see advantages and disadvantages, also in relation to the labels we drafted together?

I11  
 Well, it’s always the classic spread between what is simple and easy to see at a glance and what requires deeper digging. Of course, if I have a label where I can look at it and make a decision, that’s great. Still, I would probably always research deeper to see if it really solves the problem. Especially because these are mostly neural network things, but maybe the neural network is not what I need to solve my problem right now. There’s a bit more behind it, and whether I can really squeeze all the models I could use, all the approaches I have, into the same label with the same parameters. If we move in the direction of trust and how I interpret everything, I would almost question that and eventually still do research. But for a first overview, if I know that this label was built for a class of solutions, say neural networks, that can be evaluated in a certain way and stamped as good if it fits, and I know I need such a solution, then I can look at it and see at a glance if it’s useful. But that should come after the step where I’ve already really informed myself and looked at what I want to use. But at that step, it definitely makes sense, and I don’t have to read everything again, because honestly you don’t read all the papers in detail again. You can’t…

A1  
 You look at the one table with the relevant information or the Getting Started in the code repo and don’t dig deep into whether it really does what it claims. That’s generally a problem we have in the field right now. And what you said earlier, I think it has to be said clearly, these labels are actually meant for specific models that are practically pre-assembled, maybe pre-trained. If we now say we build a completely new system for a tabular dataset with 10,000 samples and 10 features, of course, in the end, you have to either go through AutoML or manual hyperparameter optimization, try 20 pipelines, maybe 200 pipelines, make a ranking, and then make your decision. And if you look at 2 models or 200 different models, you also can’t look at the labels; it’s not clear. Then the table actually helps more. But it’s not a form of representation meant to replace the other possibilities, it’s more complementary. Because we feel this very abstract representation is simply missing. Especially when more and more people want to use AI who don’t have domain expertise and maybe don’t even know libraries in detail, but still want to quickly decide which model might fit. For them, it’s a helpful bridge.

I11  
 Yes, definitely. If you have no feeling for it at all, which is often the case with these classic neural networks, if you’ve played around with them and know if you add one layer, then another, then another, it takes longer. If you’ve never had that experience, then it definitely helps.

A1  
 Okay, very cool. Earlier the topic of trustworthiness came up, and if I understood correctly, correct me if I’m wrong, you said that if you only have this abstract form of representation, you could rightfully ask critically, how do I know that this is actually Ground Truth, meaning that what is shown on the label is actually correct.

I11  
 Yes, referring to my problem. So I wouldn’t even imply that someone is maliciously putting something false on the label, for God’s sake. But just whether what is labeled there really fits my problem. Well, that’s kind of the...

A1  
 Yes, well, that’s a bit the problem with transfer learning. I mean, of course, the accuracy here is the accuracy on the test data, and whether it holds on the specific company dataset, you have to test it by measuring it. Usually, you can do that, unless you’re completely unsupervised and have no labels at all, then it gets difficult. Still, in the context of trustworthiness, can you think of anything that would make the label feel more trustworthy, especially regarding who might issue such a label? Like, where does the label come from?

I11  
 For me personally, if I somehow know who and what is behind it, yes, you would always want to know who and what is behind it. If there’s a university stamp on it, for example, it immediately seems much more trustworthy.

A1  
 Do you think a company would feel the same? That a university stamp seems very trustworthy? I mean, you come from a university, you know how this works. It’s all rapid prototyping. Yes, that is...

I11  
 Yes, I understand that. In case of doubt, yes, of course I know how it works, and I also know, well, never mind. I don’t even know if a company would care, like if there’s a XX XXXXXXXX stamp on it, I don’t know if a company in Leipzig would care if there’s a XX XXXXXXXX stamp on it, something like that. Yeah, no idea. Difficult.

A1  
 So we’re also asking ourselves who would be an instance that could reasonably provide such a label, because we are currently experiencing some mistrust anyway, whether in politics or in companies, vaccine skepticism and so on, which are of course huge problems that we as machine learners cannot solve. That goes more in the direction of social sciences, who probably have a bit more expertise to deal with such topics. But we are really asking ourselves, who could issue a label so that people would say, okay, now I trust the AI, because often people do not trust AI at first.

I11  
 Yes. These are just two different questions. Trust in AI in general and trust in a label regarding the model’s performance. Those are two different things, or two different trust problems, I would say. Technically, I don’t know. Because if, for example, a label from a big company, like Microsoft, Google, something like that, is on it, then everyone who is fully into open source will think, if it’s Microsoft, it can’t be good. Finding an instance that makes everyone totally happy is probably impossible. Maybe it has to assert itself as a brand, as its own brand. I don’t know.

A1  
 Yes, we don’t have an answer to that either, that’s why we’re just gathering opinions about what is relevant.

A2  
 You actually described it quite well just now. That’s how I see it for myself. It’s its own question of trust, a shift. So first the question is asked, can I trust the AI? And if you can’t answer that properly, then you ask, can I trust the label? We call that a second-wave trust assessment. And then you basically just shift it. So it goes back to standing there equally clueless and then having to check again, yes, can I trust this now? So that’s summarized quite well.

A1  
 But in itself, you have the feeling that labels can somehow be helpful as a certification to certify AI products, especially with regard to the AI Act and so on. A lot is happening there right now, and in the future we will also have liability for AI solutions, and companies will at least have to disclose their processes for how they put together and plan to use such AI products.

I11  
 I don’t know, right now, I’m not in the details, but in these matters, if a company has to disclose how an AI came to a decision, how it’s built, I fear it is so complex that a label, which states things about accuracy and power draw, is not enough to answer these questions. So if I say, I don’t trust the AI or I don’t trust the answer it gives, I don’t care how long it takes or how robust it is, roughly speaking. So I think it’s a bit different. I don’t know if the label…

A2  
 What would interest you then?

I11  
 Well, as a technical person, I would of course be much more interested in how it is built, what else is built on top of it. What we also have as an additional complexity layer is that the answer that an AI model gives is not taken purely as it is. Something is often built around it, multiple loops run, and something is added to prevent or amplify certain answers in some way. Those are all things that play a role, which I don’t think a label can really cover. And whether the hard performance parameters help with a trust question, I would somewhat doubt that, to be honest.

A1  
 Yes, so it is definitely the case that this form of label, as we have it here, does not completely fulfill these purposes. That’s the question.

I11  
 It can’t.

A1  
 Exactly, it can’t. For that, on one hand, it has to be proven under the AI Act that the model creation process is somehow comprehensible, fair, transparent, and so on. But that often already starts with data acquisition. Then it’s not about modeling at all, it’s just about how the data came about, how it was assembled, whether attention was paid to the source of the data, and so on. As it stands now, they just do it and then show a product with results, and no one cares where the data came from. And it’s slowly starting that lawyers get involved, people sue because the training data can be reconstructed, and things like privacy issues, etc. Yes, XXXXX.

A2  
 I need to ask a clarification. We also asked you what information you need for something, and we didn’t really get into the conversation that you would also need this information to build trust or use it. But we didn’t explicitly use the word trust. We talked about use or decision processes, I think. I’m just a bit surprised that it’s about this information now. And I wonder, is it because you didn’t suggest it because you know it’s technically extremely difficult to implement? Or did you not consider it in this context because we didn’t talk about trust?

I11  
 Do you mean when I asked myself what I need to make the decision,

A2  
 Yes, exactly.

I11  
 I had a completely different process in mind. If I really stand as a developer and decide technically on a model within a whole construct of processes, customer requests, etc., then the things that are labeled are important. But if I use a product where AI is involved, what matters more is what the developer who created it, or all the people and the data it was trained on, what was built around it, and how it came about. Those were two completely separate processes in my mind. So I don’t really have a problem with trust in that sense if I am the person deciding which model to use. Especially since, if I talk about building a model myself with TensorFlow, doing time series predictions or classifying images, I don’t have a trust problem. But if I write a chatbot and use a pre-made model, and I have to decide between an open solution, OpenAI, or I don’t know what Amazon or Google models are currently doing, then I have to look closely at the context in which I use it, how much I can trust the answers that come out, and what I add on top. Then I don’t trust it initially at all. I would never fully trust the answers ChatGPT gives. So these are different aspects of the whole thing.

A2  
 I’ll try to paraphrase this a bit. So, it’s a bit context-dependent which information is important. So if it’s somehow application-specific, no, not application-specific, because Chat-GPT is also a very application-specific task. So context-dependent.

I11  
 Yes, one thing is maybe also a bit about the role I’m looking at this from right now. So am I the developer sitting there, and I only have to classify data, and I have the process very much under control, because I don’t know… let’s say I build a Random Forest, that’s a model that tells me whether there is an apple or a pear in the picture or something, I don’t know. Then I don’t have a trust problem, because I can build the model myself, so I don’t have any stress with that. If I’m supposed to write a chatbot that knows the operating manual of a huge machine for a company, and I want to write a chatbot where the operator can talk to it and tell me, here, I have this problem, which screw do I need to tighten? And ChatGPT tells me screw 43, and screw 43 roughly crashes the entire machine. Then my trust in the model is, of course, much lower, and I have to build a lot more around it to verify this answer somehow and check it, that’s another technical question somehow. I actually think…

A1  
 May I… sorry, go ahead and continue first.

I11  
 No, thirdly I wanted to say, so this story, if I’m really just a user of such an AI thing, then of course I have the least trust, because I don’t know the base model and I don’t know what else has been built around it. Those are the three roles from which I look at the problem.

A1  
 I think that’s also kind of the point, of course as a developer you have the possibility to debug the model for every decision, figure out where the problem is and so on. But it’s exactly about the fact that in the end this is a model that is provided to people in the form of software that they use and maybe distrust. And the question is whether a label could perhaps bridge that, like here, I developed this, I found these important properties, I’m providing them to you now so you have some transparency about how this model behaves and you can also see, okay, I looked at different aspects as well, to build trust with the target group that is somehow behind this development process or is in focus during this development process, to reach them somehow. And to inform them. Reporting, yes. Cool. XXXXXXXXX, do you have any questions about this?

A2  
 No, I don’t have any questions, sorry.

A1  
 Perfect, then we are actually done.

A2  
 But XXXX, did you want to say something?

I11  
 No, all good for now.

A2  
 For now.

I11  
 I have to think a lot about this first. Because I find it fundamentally very, very interesting. It’s definitely a cool topic and also a very, very important topic. Fundamentally.

A1  
 Yes, thank you very much for your time and also for the feedback. We are actually now done with the content part of the study, and we can stop the recording accordingly.