Interviewer
OK, let's see. Is it starting? Yes.

Interviewee
Alright.

Interviewer
Perfect. Yeah. So the transcription is still a mess. I have to read everything over and try to hold out all the bad words, but it's a start. It's a start.

Interviewee
All right, I'll speak into my microphone properly and not use slang.

Interviewer
Yeah.

Interviewee
Okay.

Interviewer
Well, it can pick up slang. Actually. It was quite accurate with the Dutch one. You're the first English one. I'm trying, so maybe it is even better, but I'll start explaining a bit about everything, because as I mentioned, you will be anonymous. The company will be anonymous too. It's like the characteristics are maybe drawn out, without characterising too much.
And the data from this interview will be used for my thesis. So I will retrieve it. I will code it into more- Yeah. What's it called? Malleable data that I can really derive my conclusions from. And from that I'm gonna compare the interviews I have the codings I have and create a framework to basically handle accountability better in the financial sector regarding AI, if that is needed somewhere, of course there are a lot of rules and regulations in place already in a lot of companies, and that's also what I'm going to ask you about.
But I just want to see if accountability is also merged into that, basically.

Interviewee
Mm hmm.

Interviewer
And for this interview it will take around an hour. If it it takes longer, we can also plan a second meeting or see how much time we have. And then afterwards, I will send you the transcriptions when I have read them through and edited them properly, so that it seems like a nice story. I will send it to you. You can validate: 'All right. I agree that you use this for your research' and then that helps me to use it. And the second validation happens a little bit later, probably in a month and that is where you validate the further results I get from the overall data, so the framework that you maybe say, 'Oh yes I agree', or 'this is helpful' or not, stuff like that. But those are short moments. This is the longest time that I will ask of you.

Interviewee
Mm hmm.

Interviewer
The interview is divided in 4 sections and this is the first section that I introduce everything to you give you all the information. Then there are some general questions about you as an employee in your company and AI. Then the main questions are about AI, AI governance and AI accountability. Those are around 15 questions. And then we end off with some final questions regarding the future of AI and ethics.

Interviewee
Okay.

Interviewer
So without further ado, if you're ready?

Interviewee
Yeah, let's do it. I'm ready. Yeah. Yeah. I talk about this a lot at work. So I feel like I'm ready.

Interviewer
Okay, nice. So if you could just shortly introduce yourself and your function within the company.

Interviewee

Okay. So yeah, my name is [Name]. I've been at [Company] for 2 1/2 years, and that's been my time in the the the finance sector. And before that, I was a scientist in in the the Netherlands organisation for applied science in the area of cybersecurity.
In my current function at the bank I am an IT engineer for and I'm a senior engineer in the technical team of the Cyber Security Department. And my role is to manage the cloud security infrastructure, create automation and solve logging, gesturing and log engineering and data engineering challenges, through a lot of cloud technologies. We have quite at shaped team. Some members of the team are really dedicated in like cloud native technologies like infrastructure as code and terraform and things like that. But I'm more on as a Python developer so I do quite a lot of Python development. I work alongside a lot of AI teams and I've been the data engineer on a lot of like machine learning implementations throughout my career as well. So not necessarily the person who's, kind of fine tuning and choosing models or the like, but I am certainly the person making machine learning flow pipelines in Python. And then I've worked alongside kind of mathematicians and econometrics people and data scientists who who do the theory and I have kind of the role of the engineer in a lot of projects. So that's I think that's best describing where I'm at right now at the bank. But yeah, In terms of my background in studies, I studied computer networks and cyber security and I also studied ethical hacking for computer security as well. And yeah, I moved- I've slowly moved away from like a hacking kind of thing and more to, a bit more towards site cloud security because I feel that's the direction everyone's going. Yeah, that's what that's like an intro.

Interviewer
Yeah. So if you talk about Cybercloud security, what kind of what would be an example of what a responsibility is that you have or a thing that you do?

Interviewee
Yeah. So for instance, this there's different teams in most banks within the cyber security departments. Ours at [Company], it's called CISO, Corporate Information Security Office and it's a coincidence because in most organisations around the world, when you search for the acronym CISO it'll be like 'Corporate Information Security Officer'. But we also have our entire huge workforce that works in cyber security under this name and there's multiple different like, branches, but I'll just cover like a few of them. For example, there's like fraud detecting financial crime. There's like, the security incident management, which is more in terms of like, they work closely with the police and things like that. And they also look at insider threats within the bank. And then my grid, which is like a small team of around 100 people, is cyber defence. And inside cyber defence, work is split again. It's like a top down tree kind of structure and within the Cyber Defence department we manage being attacked and we manage to try and

prevent, detect, and respond. And those are what some of the team names are within this grid as well. So that they're like some of our rules and I fall in between all of these, so we've got like security monitoring, detection, prevention, response.
We have some teams that manage with like how secure our third-party vendors are and things like that. And when the bank gets attacked, there's also a group of people who work in something called the SOC which is the Security Operation Centre. Most banks have this and these are the guys that you might think of that are like on call and they might get called up in the middle of the night to really respond to an incident and to close the ticket that's been raised and the alert that's going off, that's indicating the bank is under attack and that those people might get together. And with the automations and our tools, they will use probably a lot of cloud security tools. So this, in terms of cloud security, there's different vendors of course. Usually banks should go with one. But some people use different clouds as well. Let's say the organisations I've worked at usually choose one. So you've got like Microsoft Azure, there's one, we've got Amazon Web Services, we've got Google Cloud platform, lots of different ones there. And you know when you use these cloud services, you don't have to worry about setting up new databases to store all of your security data. You just kind of pay as you ingest and the database size goes up and down and you pay a certain amount per subscriptions and business, contractual agreements with like the cloud provider. And that's the case with my current workplace and we use a lot of Microsoft features. So what one of the tools is called a SIEM and it's called a Security Information and Event Management. But sometimes it's also Security Incident, so you could technically write Security Incident slash Information, Event Management and that's really, at least on Microsoft Azure, the tools called Sentinel and that is where we write all of our rules. And we connect our rules, like for example, I can't talk about any of the real rules right now, but we can talk about like, some hypothetical ones that are like really basic in terms of cyber security. So if I said there are some some rules such as you log in somewhere and then within a second someone else logs in on your account, but they're half the world away and the IP's don't match at all. That's quite strange. And those are the kind of things that you might write as a rule like: If if this user attempts to log in like 5000 times in under a minute, it's likely someone's trying to break in big, like brute force their password, so that might be another rule that you would write and you kind of implement in the scene and there's a lot of automation that happens there. It has to be really fast as well. That's really, it's critical that some of the detections have to be like near real time and then so that's just the detection part.

Interviewer
Yeah.

Interviewee

And to do the detection, we have to get like logs, a lot of logs, from all the machines across the bank. You know some of the banks and some of these big corporations we can have like 50.000, 70.000, 100,000 machines at least. Like some the big banks in the Netherlands I estimate all have like in the tens of thousands of like virtual machines, mainframe, on premise machines, like in the data centres that we own. And then yeah, like i said, on Microsoft Azure we probably have thousands of virtual machines as well.

Interviewer
Yeah, yeah.

Interviewee
And then lots of like docket containers and it's really important to have cloud security to ingest all of that data coming off them and then we automate it all in the cloud. We connect all the data coming through this like central hub. That is the scene and that's where we also do a lot of our automation. So we might have a lot of Python there running to like fix data or to try and detect things and we use machine learning there and I'm sure we'll get more into that in a bit later. But that's kind of the idea of of what I use.

Interviewer
Yeah.

Interviewee
Cloud security for your scene. No, that's why we use it as well. And when there's an incident, it like pops off, like is detected, and it it kicks off. A ticket is made and it gets assigned to someone. So that's all managed on there as well and then that person will update the ticket with like the status like if the incident's ongoing, if he needs to escalate the incident. And in a lot of these, I can't like confirm or deny this either, but in a lot of these systems you can pre-program responses. So let's say like this is just hypothetical, but one response might be that a person's machine is locked, or we lock them out of their accounts or we cease- Or maybe we take like a forensic image of the machine where we take a copy of the hard drives. But we could- maybe we can do it remotely like at the push of a button. And the more you can take the humans' involvement away from, and the more you reduce the alerts to only real alerts and you- a big part of our job is making sure that, you know, getting rid of the false positives and false negatives and making sure that only the really important stuff gets through to the SOC. Otherwise, we'll literally keep them awake all night. Yeah.

Interviewer

Yeah. Yeah. So you want to reduce the amount of error, so that's it's not, yeah. The person gets woken just for the really bad parts basically.

Interviewee
Yeah, exactly. You know, some alerts are maybe fine to just have a look at when you have time and some alerts like might needs someone's phone to ring in the middle of the night, so you have to decide on that as well, like the severity. So I think that's like a good intro into how the cyber defence kinda works in our bank at least.

Interviewer
Yeah. Yeah, no. From your story and I also from my impression of it before, I think, I reckon it's very complicated and also hard to talk about it if you have to be secretive about some parts, but yeah. Well, just let me know if you can't answer something or hypothetical terms. That's also fine.

Interviewee
That's OK.

Interviewer
But could you explain a bit of the AI initiatives that are being taken right now at [Company]?

Interviewee
With the initiatives, I can speak about some. So we have different departments of course, like I told you and you can't just go off and use a model like will-nilly or start using machine learning. Really we have trainings that you have to go through to become a developer person, for example. And then we have trainings that you might have to do to actually use machine learning models and things like that. And for example, certain teams within the Bank of data scientists and academics- who have who have studied academically, specifically in AI and data science, they will kind of create these benchmarks. But it is really important for us all to understand the threats coming from machine learning and AI because we're in the Cyber Defence department and they'll be used against us as well, and then we have to know about it because we have to know about how machine learning and AI can be used for detection as well. So we have to be- so in our department, the literacy required is quite high. And you can tell, because as the years go on and the talk of of GPT, and generative AI, and large language models and machine learning being used to to like analyse code for secrets and stuff like that. It's a real game changer and I would say now there's been a huge push for people in cyber

security to be AI literate and machine learning literate.
In terms of the- I have a little document here on it as well. So we have some [Company] AI principles. Which I can talk about a little bit that I think you'll be interested in. I have to loosely go around this a little bit, but we have to be (1) lawful. Of course, that's really important. So we have to respect all applicable laws and you know, we're under lots of different regulations and laws. We've got GDPR. just like everyone else, and there's an upcoming European AI regulation and we don't want to be caught out by that either. So we, well, I've heard that the data scientists at [Company] are closely monitoring how the European AI regulation would affect us. Because obviously we don't want to invest massively in AI and then be kind of shut down if we put lots of man hours in a projects and we find out that it's likely that European AI regulation won't allow some things. And then yeah, in terms of the the Dutch laws and regulation, we of course respect them and we try and we have legal teams to reference there. And then yeah, apart from lawful we have. What is it? (2) reliable. So we need to assume that everything that we do use is going to work as expected. I don't know if you've heard the story about Canada Air, but their AI chat bot. They didn't- they tried to fine tune it to answer questions from a vectorized like rug database, and basically it was supposed to be an assistance bot for the airline. So people, so someone asked it if there was compensation if you'd recently had a loved one die, if they would like compensate the price of some of your flight. And what happened was-

Interviewer
Hmm.

Interviewee
It lied and said yes, but that was not the case and it kind of hallucinated that and that was not in the database, but because we're starting to kind of normalise using models that based on our data and pre-built models. This can happen, and that's what I feel like the reliable clause is there for. Because if we willy-nilly use big LLM's that have been trained on the Internet, and then we expect our customers to ask it something about a mortgage or something like that, if it gave a completely hallucinated answer that was not the case for the bank, perhaps we could get sued or we could get taken to court and we could- They could say: 'Your software told me to invest' or 'your software told me to do this'. So yeah, it has to work the way we expect it to. There and then. Yeah. The final one is (3) ethical. I think perhaps there's more. So I can't say like this is- I can't say like in terms of my bank that these are all of them. These are just the ones I kind of remember, to be honest. But in terms of ethical, we have to like take into account the rights and interests, especially the rights of all the stakeholders. So if it's colleagues that are going to be using the program, we have to think about, like, not, you know, the ethical, the ethical rights of that in cyber security, before I started at the bank, I researched the

ethics of phishing. With phishing, when you send like a malicious e-mail or like a fraudulent e-mail to someone, it's a huge method of attack in cybersecurity. It's one of the most popular, like initial initial attack vectors. And the thing about it is that you can kind of train your staff on it through various methods. You can do trainings, you can teach them about it, you can hold workshops, and you can also host simulated phishing attacks and tests. So all of the employees of a company will be sent a phishing e-mail and if they click on it- It's kind of made to look malicious. We'll send it from an address that they shouldn't have clicked on; an unknown, one-time-use address saying that they have, you know, we'll use social engineering techniques like you only have X amount of time to click on this link and they'll click on it and then we would do like a training.

Interviewer
Oh, Okay.

Interviewee
Uh, they'd be prompted to do a training. Or maybe their manager, their manager would have to speak to them or, you know, I can't say what we did because I just can't say at all like what we do at the bank in regards to this. But I researched this at one point and and a lot of people get upset and anxiety ridden about when they have to tell their co-workers they're going on a training because they clicked on it. Phishing e-mail people get really upset and people don't like it. But.

Interviewer
Yeah, you could basically called out for your weaknesses.

Interviewee
Yeah, uh huh. In America, there's even been firings with people failing more proficient tests and people losing their job, and yeah. But at the same time, it's been proven as very, very good because people are scared.

Interviewer
Yeah, exactly.

Interviewee
Because people are then scared, not only because of like the reputation damage and the cost, financial damage and the cyber attack risk, but also because they don't want to be the one that did it. They don't want to- they don't want to be the cause. Yeah.

Interviewer
No, no. There's also a sense of maybe, yeah, a feeling of responsibility for your workplace that you don't want things to go wrong and stuff. Yeah, but-

Interviewee
And just the reason I brought that up was when we were talking about like ethics, just because they're like- the stakeholders aren't clients and we still have to consider the ethics of, like, how we'd also treat like internal staff members. So there's a lot to think about in terms of like ethics. There's like fairness, well-being. Trying to remember the other one, like transparency I think is really important as well. Like at the start of this meeting we already talked about like how the meeting would be used, what it would be used for, both of our backgrounds. How we both feel about what?
How how we both feel about about it. And I think when you start to use people's data and you start to use people's input data as well, that they can input into your like product that might use AI or their applications giving you data, you have to be really transparent about how you're going to use it, how long you're going to keep it for, like in terms of retention, how long you might keep the data for, where you keep it, if you're going to delete it after a certain time, who's going to have access to it? These are all really important things that apply also to AI, think.

Interviewer
Yeah. Yeah. And so. Let's see, like the so, for instance, within the company, do you see a lot of AI being implemented? Or do you see a lot of potential for it? Like does it feel to you as something that is being like caught up on or is it more like, oh, they're really trying to innovate in that?

Interviewee
We've we've been doing it a long time. You know, I would say that. I can't see certain things here. I sorry. So the bank has machine learning teams and AI teams separate. An they are- We could definitely always do with more. It feels like they're in need everywhere across the bank. I think I can see, and in my previous jobs as well, there really- it's a really valuable skill that can be applied to a lot of different departments now. Previously, maybe it was only being used for, you know, insights, business insights, trading insights, things like that. But now it can be used everywhere. One of the public use cases that we talk about, that we have partnership with Microsoft on and OpenAI, is we that they talked about the Microsoft build event at Schiphol recently. And when they did a workshop on generative AI is that we use LLM's and GBT to summarise conversations with customers. So customers call and essentially you'll be able to find

some information about this online as well, which could be useful just to make sure the core is is correct. It's that the conversations are transcribed and the notes that are collected, or there's a prompt that will kind of sort the transcription of the of the whole thing into a nice, like organised section, section prompt. So maybe it'll have like the action points that need to be taken. Maybe it'll have like what was discussed. Maybe it'll have like, I don't know, in the future. I don't know if it does right now. This is kind of speculation, but maybe it'll have like a sentiment of the meeting whether it was like positive or negative, or angry or happy or sad or something like that.

Interviewer
So is it like the Microsoft Copilot that you are talking about?

Interviewee
So Copilot is more to to do with like assisting everyday work.

Interviewer
Yeah.

Interviewee
But this is like more the GPT OpenAI as a service so that this is like AI and it's the LLM's, like large language models, and then the bank has built some automation to do it. The bank has built something to automate the process.
There's a lot of different programmes that people are using right now. There's an open source one called like Landchain and there's another one that Microsoft's building called like Semantic kernel and we're investigating all of these a lot right now. Loads of teams are investigating by themselves and then some of our big teams to do with AI make the final decisions, you know, and these kinds of decisions that the bank will be like, the platforms that we use, like, oh, we'll use, like, data bricks. For example, what we we'll use Microsoft Open AI service and, yeah, I think that's for in cybersecurity, I can tell you that we use it, ml. We use machine learning. But in terms of AI and generative AI, we're moving towards it quite fast. And there was a recent quote, it was saying that the US has implemented a lot of AI and they have very little regulation, and in comparison to that, I'm not saying we have loads in the EU, but in comparison to the US, in Europe, I believe that we have a good amount of regulation coming and in the pipeline and already existing. But the amount of AI we've implemented in industries is fairly low compared to the US.

Interviewer
Yeah.

Interviewee
So we- I'd say, if you like, do it geographically, you're kind of slightly behind, but we're gonna have probably stronger regulation which is gonna be good. I think the sentiment of regulation is good and especially at the bank, no one wants to run off by themselves and mess it up for everyone else and kind of ruin it for everyone. So yeah, we tread carefully there especially when it comes to clients, you know. I'm not gonna say if this is the case at the bank. But if it comes to certain like- if it comes to just trying to get some more insights from a static data set, it's not really going to hurt anyone using LLM's on that, unless it makes a huge- unless the data includes ethically risky things like personnel identifier- PII (personally Identifiable Information). And if it includes, you know, things like, where people are from.

Interviewer
Yeah.

Interviewee
We all know about what happened with the Netherlands' tax authority and things like that. So yeah, so that's kind of the area we make decisions based on the data, as well on how ethically risky something is. And I think we're being really careful at the moment, but it's quite hard for me to answer because I'm not involved in the decision making. I'm just, I'm just, I'm more of a user, like not a user but a developer.

Interviewer
Yeah. Yeah. So you basically witness it, but you don't necessarily- You're not involved in the, as you said, decision making or stuff like that?

Interviewee
Not in the policy, not in, not in terms of policy. Like I can tell you that I've tried to ask if I can use it for a use case and the use case has been like; I had to write several documents about what would be used, how, where, and all these things, and then it was scrutinised by a team to decide if they wanted to use it in a pilot. And there's a lot when it comes to LLM's. I would say we're very much in the pilot stage.
We're trying out different things. We're trying our Copilot for existence and seeing how it goes. Making a decision as the bank whether we want to use it or not. And then we also

have pilots for actually like implementing it in business processes as well. And that's why we're a bit different. So yeah,

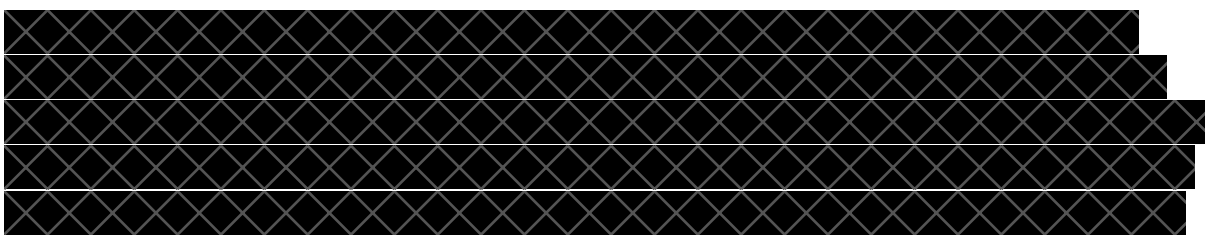
Interviewer
Yeah.

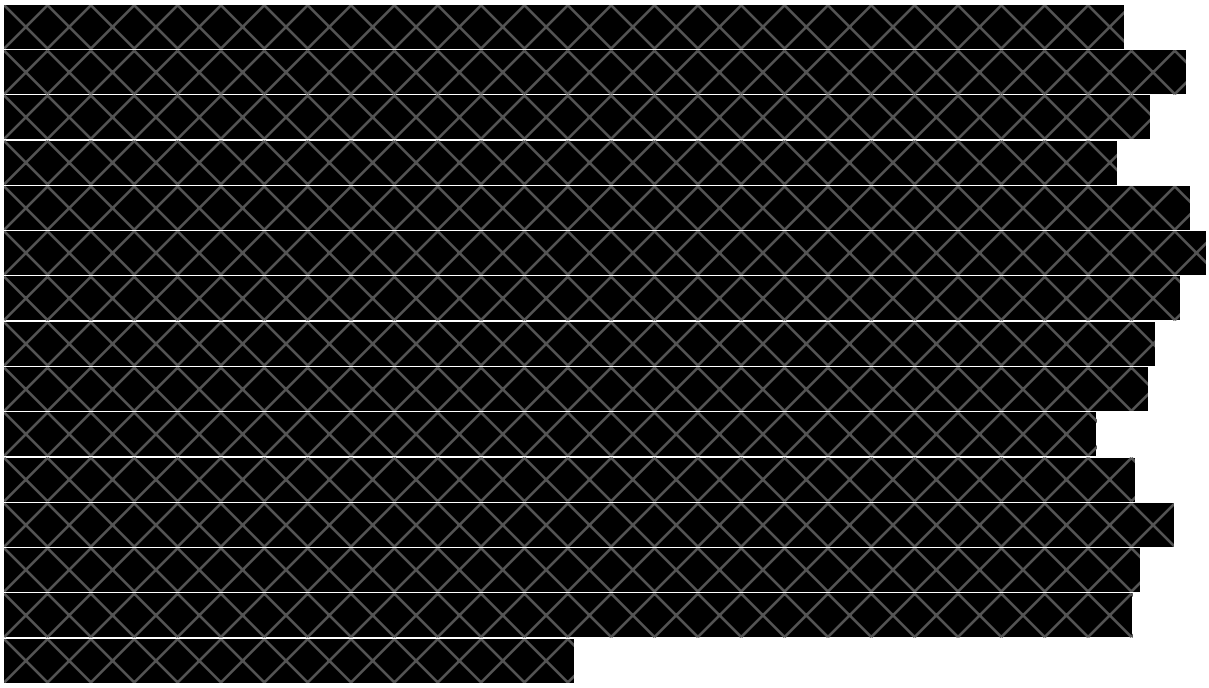
Interviewee
I've kind of mixed it a little bit up in this last section of the interview. But machine learning has been around a bit longer than AI, right? And I don't know in most universities these days, when they get you to do something with machine learning initially, maybe they have you do like classification or clustering or something like that. And those techniques work really well in cybersecurity, like classification anomaly detection. These are all things we use a lot and like we talked about before, if you have a log of like people entering their passwords, and the anomaly detection attack detects like this one user that 5050 logins are happening every minute like that's, yeah.

Interviewer
Yeah, it can do like a probability calculation of like I have these factors, these ones coincide with something that is- something fishy is going on and I can decide from that, yeah.

Interviewee
Mm hmm.
Exactly. There's some things where maybe you don't need them. You can just use a simple count like, oh, if over 1000 logins happened, please detect this. But there's other things more like advanced types of attacks on banks, where machine learning can really help identify patterns that a human might miss, or identify patterns that a human does not have time to look at all the time.

Interviewer
Yeah.

Interviewee




Interviewer
Yeah.

Interviewee
'Cause my team manages a lot of the data engineering for the Cyber Security department and then it arrives in the tables and gets used.

Interviewer
Yeah. So it's a lot of like cleaning, normalising that kind of stuff is what you would use it for to make the process faster and easier for you, yeah.

Interviewee
And if you think about it, we can't do it on every log line because that would cost enormous amounts of money because as you know, you pay.

Interviewer
Yeah.

Interviewee
You pay for using the OpenAI service, for example. And machine learning as a service, we have different levels. We have like infrastructure, service platform as a service, software as a service, you know, let's say you like pay for PayPal premium and it's just

like a- you pay for like LastPass, like password manager, or something like that, or Spotify. That's software as a service, like you don't have anything to do with running it. You don't have- You can't like launch your own programmes on Spotify or anything like that. So that's like a software as a service and then you know you've got platform and infrastructure as a service and on platform as a service, maybe you can spin up some VM's (Virtual Machines) and things like that. Sorry, on infrastructure services, you can spin up some VMS and on platform services, you can spin up some like applications. So there's there's different like layers and it's the same with machine learning and AI. That's the same level, like you can choose to just buy some API keys and write all of the code yourself, host it on your own infrastructure, pay for all that part, but you're still going to pay. I believe for the amount of text tokens, when it comes to LLM's or when it comes to training a normal model on machine learning, you're going to pay for maybe the amount of hours of graphics cards and GPUs that they had to put in to to create your model or to fine tune your model. And these costs are- could be crazy as well.
So when it comes to my job, with millions of lines of logs, I can't send them all, so I have to- So that's where clustering comes in really a lot as well. And clustering is a really difficult problem and we keep an eye on machine learning and AI a lot for new solutions, because it's moving fast. In the last few years the way computers understand human written information is like going quite quick and I think we'll get to a point where it's going to be a problem that's solved. Like, for instance, if you pay five engineers to look at some logs and to normalise them, it could take quite a long time. And I would say we're still at a point where most organisations in the world often manually normalise, manually cluster. But I think that pretty soon that's not going to be the case. But what we've seen in the last year, GPT and LLMS and things like that.

Interviewer
Yeah, it's going quite quick, yeah.

Interviewee
Yeah.

Interviewer
So let's see, I'll ask you a few questions.
Because let's see. Okay.

Interviewee
So right, we can go- We can- I can give shorter answers as well if you want.

Interviewer
No, like all the data is data, so it's nice.

Interviewee
Yeah. Okay, good.

Interviewer
But I think we'll run out bit out of time. So let me know if you have to leave or something. I think I can run out like 15 minutes, but somebody's cooking, so I don't know when they are done.
But I would like to know like a bit more structurally, what is your definition of AI?

Interviewee
Yeah, it's a tricky one. It's a tricky one that I don't want to get wrong.
I think there's a there's a definite gap between what most- There's different levels of thinking, like almost. There's first understanding the difference between machine learning and AI and there's also understanding premade models, right? I would say my definition of AI is being able to give it challenges, and where it works out the solution itself, but maybe it was not specifically- Maybe it was not- it wasn't made for that exact problem. It's quite dynamic and it comes to the solution itself through solutions like neural networks and things like that, where it it comes to the best solution. I can also like word it like you can fine tune a large language model. You can like- you could like- there's a programming language to make microchips. It's pretty wild, believe it or not, it's been around a really long time and like when you get to like the fundamentals of computing.

Interviewer
Okay.

Interviewee
It's like a bunch of transistors, like logic gates and ones and zeros, and there's a programming language where you can write what you want and it translates into, at the end when you've made it, it will literally give you like this picture of like lots and lots of micro logic gates and transistors layers, and it will automatically like, make layers and layers and layers, and then what you've got right there is like the hardware descriptor file and every log to make a microchip. And we know that GitHub worked with one of the biggest microchip carriers in the world, AMD and they fine-tuned an LLM to specifically

work with- to specifically be good at assisting in writing this code which is for a vary log. And that is- just because they fine-tuned it, doesn't mean it's not AI anymore.

Interviewer
Yeah.

Interviewee
But it's because it's still using the huge like large language model, but it's it's also not what everyone's defining is like generative AI, so that like now that- ever since I learned what generative AI was, because I'm also not a machine learning and AI expert, then my perspective of AI has changed again. And I'm like oh, we're not, we're not actually there yet. We're not actually at AI like what- I grew up thinking of AI with like Terminator. Where you know the machine's making its own decisions, and maybe it turns on us one day. Or maybe we-

Interviewer
Yeah, yeah.

Interviewee
You know at Google, Google have this, Google and Microsoft and everyone joined up in the US and created this like governing body because we don't want to make- we don't want to step too far with AI. And you know it gets quite sci-fi quite quickly, quite creepy, but it's a real concept like a bit like kind of rolling off the rails. It's something I don't know a lot about, but in terms of that they are my 3 levels right now; machine learning, like we talked about like classification, clustering, anomaly detection. Then when it comes to AI, you've got all of these tools where you can like dally- where you can give descriptions of images and then it's actually like generating content. But it's based on transformers and it's based on technology that's trained on huge data sets and like billions of billions of like like parameters and transformers. So in the-

Interviewer
Do you use it yourself a lot?

Interviewee
Yeah, I I have a private subscription for for open AI and I use that for everyday things and then I'm in some pilots at work to use it for code assistance.

Interviewer
Hmm, Okay.

Interviewee
But for my private code, I use Copilot and it's a bit- it's it's great in that sense because you know it's it's reading your code and giving you suggestions but you know it commonly makes mistakes. It commonly doesn't understand the context of your entire code base. But it in six months, it's went from really, really bad to to really to really good.

Interviewer
Ha. Yes.

Interviewee
And it keeps getting better as well. But there's this thing inside if you don't use it, you'll lose it. So you don't have to become too too happy with Copilot, otherwise suddenly you'll be a bad programmer and you won't be- you won't write clean code, like you can't. You'll write garbage code and just expect it to clean it up for you, which is not good. And imagine, like, if we go all the way back to the start and we talk about being ethically responsible and stuff.

Interviewer
No, exactly.

Interviewee
Like how can you be ethically responsible if you don't really know what you coded and you just gave a description of what you wanted to code and it automatically filled it out for you. Yeah. And then it's like, oh, yeah.

Interviewer
Yeah, it becomes basically, it becomes too easy. Perhaps that it's- You still have to stay on top of it, but we can touch on that later as well, I think.
But you just mentioned a bit that you were working on pilots with AI. And do you have an active role in then the also the implementation process of this or is it just working on pilots and then giving back the what came from it?

Interviewee

Yeah, it's the second part. It's writing feedback quite frequently on how it affected a lot of things, like obviously productivity, how much I enjoyed it, if it made me like coding more for example, and then how well I think it performed, how much time it saved me is obviously quite important to the bank, yeah. These are all things I would answer quite frequently. And then we talk about what we think it's applicable for. We have like a hackathons.

Interviewer
Oh.

Interviewee
Pardon me, we have hackathons as well.

Interviewer
So you tried to hack the system they are trying to pilot.

Interviewee
Oh, hackathons. I know that it has the word hack in it, but they're more like brainstorms and they're kinda like a bunch of developers. They'll get together and sit in a room and eat pizza and they'll try and solve something like for example, there's a- They'll find like an LLM that can make a song out of anything and they'll decide to like- I don't know. Give it words from, I don't know, like- Trying like the Formula One recent news story or something like that. So they'll make a song using machine learning out of just, you know, useless things to test the technology and what it can do. But for fun in that sense. And then we have serious hackathons as well where we will see- 'cause you know, you need to explore these new technologies and the bank's not ready to commit immediately and say like yeah, let's use LLM's to solve this problem. Someone has to try it first. We have to have research spikes. We have to have these like hackathons where things are tried and if it goes well, then maybe we present it to colleagues in the bank when we're sure that we think it's a good idea. And then we go through all the processes we talked about, where you would start to make solution designs, solution intents, go to approval boards, speak to legal approval boards, speak to AI and oversight committees on machine learning and AI and get it all approved, and maybe Ethics Committee would go as well if it's like using customer data. Especially, I'm not saying those are the exact steps for [Company] but those are like common in the financial industry. Yeah.

Interviewer
Yeah, that- so that is like the whole implementation process of a new model or system?

You have to go first, check it out and test it, and pilot it and then pitch it basically go to all the boards, show that it's a good idea, show the numbers and then perhaps it gets implemented.

Interviewee
Yeah, and if there's like 100 teams asking for something and the bank doesn't have it, from my experience, they'll start to investigate it on on making a final decision, because even because they want to be able to say no or they want to be able to say, yeah, we have it. So and that that is sometimes the way it goes as well, they won't be peer pressured into allowing you to do it though. It'll stay off limits until the bank is ready to make a decision. So that's either like using it or that can go for like asking for something new. Like if Microsoft released a new- a nice online portal where you can kind of use different agents, like have one agent that plans the routes on a holiday, have one agent that like plans the finances, have another LLM Agent that like, you know you've got a lot of these new technologies coming out that use multiple LLM's and this is all new technology for example. I know that someone in the bank in those teams is looking at how we could use that at [Company] and also they're trying to decide who we go with. If we go open source and we do it ourselves, or if we pay just for some API keys, or if we pay for like when we talked earlier about like software as a service and platform service, if we just pay for like Microsoft. Or if you pay for the day breaks because you know we've got options. The technology's moving really fast and yeah.

Interviewer
Options, and decisions to be made basically.

Interviewee
Yeah, this has been called vendor lock; if you like build your entire organisation around one vendor like Microsoft or Amazon, then later if another vendor seems to be doing a lot better or has much better technologies but your company's so focused on only working well with one, it's really hard to implement the stuff from the other vendor. So ideally you stay quite hybrid quite like vendor agnostic, we call it where you could use things from different vendors.

Interviewer
Yeah.

Interviewee

But that is an entire section of cloud computing that we don't have time for in this meeting, but it affects machine learning and AI, so we can stop there for that one, yeah.

Interviewer
Yeah. And what is good performance mean for AI systems, in your opinion or if you?

Interviewee
Not overfitted, you know not overfitting: Huge problem. Like 'oh yeah, my model has true positive ratio-'. Oh, sorry. No, it's always the accuracy. 'Has a really good accuracy. Really good accuracy'. But maybe it's also flooded with false positives and then it's overfitting. Do you know what I mean? Like. Oh yeah, it did. It successfully detected these attacks, but it also detected a lot of benign stuff that was not an attack and that's not good. You don't want that. We want just the attacks, please. Because, like I told you in the beginning, you know, we got real people that are gonna investigate this stuff. And if accuracy is low as well, like it becomes closer to random.

Interviewer
Yeah, yeah. So the accuracy.

Interviewee
And we have to decide where it's worth it for us to do it.

Interviewer
Okay, so there's like a threshold of accuracy precision?

Interviewee
Yeah.

Interviewer
Everything to make sure that the system works as accurately as possible.

Interviewee
So, you know, we'll look at accuracy, precision. We'll look at the truth matrix and truth table and we'll try and make a decision whether we need- because you know we can make systems of human in the loop. Where we get a human to check it and then we retrain and it gets better. So sometimes we have to make decisions if we think it's worth it for now to run it with low accuracy, and have a human in the loop to try and add to the

training database. Like if you add more ground truth and more answers, then when you keep adding them to the training set and like retraining with traditional ML, or if you add it to the RAG database in LLM's, the vectorized RAG database. But it gets bad over time, so sometimes you might even have to make the call if it's worth investing in a long term strategy like that, or if you're just going to hire 100 engineers for the summer and solve the problem. That happens in all industries, as migrations and like huge challenges that ML and AI could really help with, that we currently maybe hire contractors for.

Interviewer
Yeah.

Interviewee
Well, ideally the bank would. Well, banks, all it institutions would love to be able to just solve the problem automated with ML. But yeah.

Interviewer
Where do you think accountability comes into play? So, for instance, if you look at a system that has true positives, false positives, false negatives, and something goes wrong. Will there always be a human in the loop to recognise these falls and instances?

Interviewee
I would hope that the risk analysis and the process that we went through, has already made it so that it was decided that a certain system wouldn't be allowed to have the low certain thresholds or be able to make such a catastrophic mistake. I would expect that you wouldn't get to go ahead for certain use cases, unless it was known that it would perform at a certain level.

Interviewer
Yeah.

Interviewee
Because accountability and transparency is obviously really important and integrating the principles of this is again my job. Not like, not exactly creating the policy, but as a developer integrating these principles is the the difficult thing. Like for example, we have data sharing agreements, we've got like guidelines for derived data. What else do we have we have? Mandatory data learning, perhaps for developers, and then we have all of these checks before you're like allowed to use your use case. And in terms of accountability? If you've got 60% accuracy and then you're like using the results for something, and you're not- don't have a human in the loop. It depends on what your use

case is. If it's just like to make a slightly smaller batch of something that is already huge and you just want to like decrease the batch a small amount, but all of that will still be analysed manually, then I think it doesn't matter as much. But if there's a possibility in the small 10% of stuff you threw away that there was something important in there, then you're gonna have to be accountable on some data you missed. Then that's not okay, that- then you have to make a strategic decision.

Interviewer
No.

Interviewee
Because you don't want to throw away data that's important. For example in, yeah.

Interviewer
But are you able to or is that something that's very like controlled?

Interviewee
It's a difficult one. It it really depends on the use case and the stakeholders and all the factors we talked about. It really depends. You do have to decide, especially in my field, what data are you going to throw away? Are these logs useless or are they just like, every time a system connects, which is like every second, do you want to keep that? Is it useful for cyber security incidents? If not, throw it away. But if AI decides, it makes that decision for me and accidentally throws away important logs, and then that system suffers a cyber attack, who is ultimately accountable? It could be me as one of the developers who was accountable, I would say. And then the bank could suffer, you know, reputational damage and damage in another sense. And that is all a part of risk. And it's a reason why we have risk frameworks and cyber security frameworks and lots of these processes. And it's all about prevention and that sense, right? And then governance. Yeah, exactly.

Interviewer
And probably yeah, yeah, yeah. So that flows nicely into the next question. What does the term AI governance mean for you?

Interviewee
Hmm. What does it mean for me? Yeah. Not to like repeat the same things but it means- the correct- So I think- I wanna give a a good answer.
In terms of in terms of governance, like following the guidelines and following the

policies and rules set out for me as a developer. It sounds like research for sure, like prior research. The fact that there is governance means that I probably have to prepare some documents and then that will prompt me to go and read about the policies and governance, about what I have to adhere to, about the ethics and the accountability and I will have to explain to a governing body within the organisation about my use case and they will ask questions. Governance in other applications that onboarded that haven't been Ml you know, they cover stuff like authentication and like who's going to be able to access it? Who might it affect if something goes wrong? Reliability. What happens if it stops working? How badly are we affected? Do you have a like exit plan for if the provider doesn't continue, like goes bankrupt or something? And I feel like AI governance is very similar. But it covers the more specific parts, I can't answer it very well unfortunately, because I I am just an engineer, like one of the implementers, yeah.

Interviewer
No, that's good. It's uh, still something I can take with me.
I think we discussed it a little bit before as well. Can you describe some possible risks that are connected to the AI initiatives in [Company]? Yeah, possible risks.

Interviewee
Possible risks- I can discuss possible risks. I can discuss possible risks for the financial industry in general, yeah.

Interviewer
Yeah, yeah, yeah, that's good.

Interviewee
So for example, this is complete speculation, but we can- we have risks through the use and we have risks of it being used against us. There's a lot of talk of it being used to, for example, create customised phishing emails, let's say like maybe it would take someone who- maybe someone saw you and you're a millionaire. Typically you're going to get some like advanced attacks happening on you. You're going to have people calling you, pretending to be your tour manager if you're like a celebrity. You're gonna have people like telling you they're from the bank all the time. You know, people like calling up your phone provider and trying to pretend to be you all the time, and yeah, that takes time for an attacker to do. And then one of the crazy things is that now we think that using things like GPT and LLMS, it's really easy to make really, really realistic social engineering attacks like through something called neuro-linguistic programming. Because it can scrape the Internet for all of your information and it can write something really convincing in an e-mail.

Interviewer
Yeah.

Interviewee
Whether it's like, sorry, maybe it finds the names of your brothers and sisters online, it sees how old you are. It sees what country you're from, what town you're living in right now. I'm sorry. I'll just go grab the doorbell. But, yeah, maybe it's that's how it's used in attacks, one second.

Interviewee
So yeah, in terms of risks from AI at [Company] right? Well, in the financial industry.

Interviewer
Yeah, yeah.

Interviewee
So that was one risk. It's a bit more on how it can be used against us, but then the other risks are: Overuse I think, becoming too complacent with replacing humans with with with automated processes has always been a concern, like through the last 100 years, never mind just recently and you know, people were really scared when-

Interviewer
Oh yeah.

Interviewee
You know, I don't if you've ever seen Willy Wonka, but the person who is screwing on the toothpaste caps got replaced by a machine and now they're also scared about GPT taking all the jobs. But yeah, people are scared of of being replaced, but also some of it is actually important and makes a difference. Like you know, like we talked about before with the the phone calls and stuff like that as a use case. You wouldn't want to have like a GPT therapist, would you?

Interviewer
No.

Interviewee

You wouldn't really want to have like an AI counsellor or HR representative. Well, you go to them upset and you get sent to a chat bot instead. And those are risks for sure, it can get very frustrating when there's too many chatbots. I'm not gonna lie, it's already been affecting me, yeah.

Interviewer
Yeah, the human side gets taken off basically.

Interviewee
Yeah, absolutely. Even when you need support for something IT related, when you speak a bit fast or it stops to understand the context. Let's say like, you've described a problem and then the problem gets more and more in depth and then it like forgets about the initial context of the problem and it just starts focusing on the most recent part. Like I'm not saying- I don't know. Different models behave different ways, you know, like the new GPT 4 was very fast. It takes less time. It puts less effort. It's a closed box model as well, so that's a risk as well. We don't know specifically unless the owner tells us, the creator tells us, unless OpenAI tells us how it works and for the new GBT 4, which is just about to come out, they haven't told us specifically how it works. So we have to make a decision on like, do we use this model, which takes a bit longer, but it double checks things are correct? Or do we use this model which is really fast and everyone's gonna love it 'cause you're gonna be able to speak to it in real time, but maybe it hallucinates and makes stuff up and gives, really-

Interviewer
Yeah, you have to just take that as a given.

Interviewee
Yeah. And those are risks for sure.

Interviewer
Yeah. And what's- so we talked about it a bit before that there are a lot of boards and a lot of lists and frameworks and rules that you have to go through. Do you also see them mitigating these risks within the company? Well, maybe not specifically the ones you talk about, but-

Interviewee
I would say by making like rules, yes. For example, they they make rules like data steward and like data custodian. I think they even have like- I don't how much I'm

allowed to talk about this, but there's like learnings of course, like about how to do data science at [Company]. But then there's also learnings on privacy and also like data issue management, data quality control and you know that stuff's existed for a long time. And I believe a lot of it needs to just be slightly altered for it to include AI, that it doesn't necessarily need to be brand new courses from what I've seen. Because we've already had a lot of ehm, like the learning data- what's it called?- the data learning paths have always included things like privacy, data management, life cycle management of data, issue management, data quality management. And a lot of it does apply for AI, but I feel like a lot of it needs to be adapted as well. So I feel like that's how they're mitigating it. They're adapting what exists right now and kind of their modular processes to include all the AI things. So like maybe we already, like I told you, had cloud approval boards and things like that to decide if an application should go online or not, now if that application includes AI; additional steps? Are there additional stop gaps, checks, and governance, and review boards, and ethics boards, and legal, legal consultation and things like that? Maybe you literally have to get signed off by one of these rules I was telling you, like a data steward, to make sure that the data that you're using is quality.

Interviewer
Yeah, yeah.

Interviewee
Yeah, but in terms of mitigation, in terms of- I would say the roll out process is also the mitigation. Like if they just released it without any controls or any research for anyone to use, it wouldn't have been good. They actually block a lot. Like for example, we all had Edge browser and then Copilot suddenly became enabled on Edge browser, like Microsoft has added it one day. And we had to disable it immediately before people started like putting company data into that because you know then, with company data we need to agree with Microsoft that it's going to be private to us.

Interviewer
Yeah.

Interviewee
So if people start like putting all company data in all these GPT models, that's really not good. And to be yeah, no.

Interviewer

Yeah, but they asked you to- So they ask you personally, please turn it off and do not use it, or can they control it?

Interviewee
Yes, they turn off.

Interviewer
Oh, they turn it off, okay.

Interviewee
They can control it. They'll block websites, they'll block features of Microsoft websites and browsers, they'll block certain computer programmes. But then one of the really interesting things I think you should take away, is that I believe from my interactions in workplaces that there's a severe risk, if you do not implement some kind of Copilot chat or GPT chat or assistance, that people are a) gonna look elsewhere for employment opportunities. Or b) are going to illegally, and you're gonna have a huge risk. And I can say this as a cyber security professional. Not saying that it's happening in my organisation.

Interviewer
Yeah.

Interviewee
And you know, maybe if you worked in security, you would even want to watch out for this as well, but people copying and pasting company data, data exfiltration, data privacy, and we don't want people taking data. You have data protection to stop people already sending data out of the company. And now what we have to worry about is people putting it into models, and GPT, and copy and pasting company data into these models that could potentially be used for training of the model and get lost inside of these training sets and maybe someone will be able to reference like our- I think like, at one point you were able to ask GPT if something was written by GPT.

Interviewer
Hmm.

Interviewee
It's changed slightly now, but that kind of shows that it has reference of what it's asked.

Because remember that you're making API calls off to the model, which is hosted remotely, and in most cases it's not like a static model on your system that's not being retrained. So [Company] will also, one of the things I haven't really mentioned yet, they will make huge agreements to make sure that no data that is entered officially is ever used or is kept or used for retraining. It doesn't go into retraining of the model. Yeah.

Interviewer
Yeah. So, okay. So they're also- they're trying to keep on top of their AI governance a lot, or data governance in general, and also actively are trying to evaluate this governance as well to make sure that it's airtight and the data doesn't go anywhere?

Interviewee
Yeah, exactly. Because if they just enable GBT private and they let people on the website, people are gonna be asking it to help them, right? Confidential documents and all this stuff. And then I think personally, I think I can talk about this a little bit, by enabling it in Copilot and making pilots and saying that the rollout is on its way, it decreases the risk of people illegally and against company policy using these products, which could potentially cause us huge cyber security risk, information risk and things like that. And I think across the world it's a huge issue, that right now. And that is why we've seen spending in AI times by 6 or something, Ithink I read this morning.

Interviewer
It's very it's booming even more than- Yeah, it's growing exponentially in my opinion.

Interviewee
Yeah, and we, we. Yeah.

Interviewer
Yeah, but it's also interesting what you mentioned about, like if you don't implement it, people are gonna look for it in another way, like either another job or illegally. And do you think that is something that can be controlled by an organisation? For instance at [Company], as far as I know, they don't have a GPT internally yet. Do you reckon that that is something that can be controlled among the employees or not?

Interviewee
I think that we will get one personally, the bank could make a decision tomorrow saying no, but I hope that it will get enabled. And from their initial pilot programmes, and how they're talking about it, and how they're doing hackathons, and their attitude towards it,

their careful attitude, and like processes, I feel like I'm quite proud to say that at least I think we will have a good implementation of it that is not rushed, that has all the governance there ready, is in line with current EU and country regulations and that we're going to do it the right way. But I can imagine that people are- other companies are too small to do something like that kind of large governance process and they will just enable GPT for their employees. And maybe they'll come- maybe they'll use Microsoft, which I think is about to say that they don't use any data input for retraining. I think at the Microsoft event it was said five times. People asked that question so much like: 'will our data be used for retraining models?' 'Will our data be used for retraining models?' and this they have to keep reiterating like' no, it won't'. 'It won't be used', 'it won't be used for training, again models', but some some organisations and some platforms might reuse people's data and that's why we've talked about the transparency as well before. So [Company] wants to be really sure of that as well.

Interviewer
Yeah. And have there been-

Interviewee
Sorry, I'm not allowed to say [Company] will be but I think financial institutions will be.

Interviewer
Okay. And have there been recent changes to the AI governance structures in the company? Have you seen that for instance, you mentioned that a few like laws like the AI Act, something that's quite recent.

Interviewee
I mean. The data management learnings all were changed in the last few months. And that means that people have to do new courses, and get new certificates and things, and attend new workshops and trainings. When they do something like that, so you know it's not just articles online that you might have to read. It's like your job role has changed.

Interviewer
Oh.

Interviewee
You have six months to complete this, or your job is in jeopardy. Maybe. I don't know. Yeah. So it's pretty serious when they do something like that. And then the governance,

like we talked about. We have like awareness campaigns internally of being attacked using AI. That's not really governance. I think we we're keeping an eye on the the AI Act and the new EU regulations. And then in terms of governing, I think they have like this thing at the start, which is like the lawful part, reliable part, and the ethically responsible part and they're like the AI principles. And then we have these huge change risk assessments, so if someone wanted to change an application, to use AI, there's no getting around it. They've made sure that in these new change risk assessments, if you change to use something that has AI, you also start the AI principles and the management frameworks that would assess you on the advancing lakes and AI governance. But sorry, that's, I don't know more than that, yeah.

Interviewer
Yeah, that's no problem.

Interviewee
Okay.

Interviewer
I think it might be good to stop here. I do have a few questions left, but I think we will definitely run out of time. So would it be alright for you to maybe have a second meeting? That's definitely shorter than this one, but.

Interviewee
Yeah, that's fine.

Interviewer
Then we can just not rush anything or something.

Interviewee
Sure.

PART 2 2024-05-23

Interviewee
Oh, it started. Yep.

Interviewer
Yeah.
Right.
Yeah. OK.
We stopped at the third section. So we covered AI, we covered AI governance and now it's mostly about AI accountability. So this section also starts with a question. What is your definition of accountability?

Interviewee
My definition of AI accountability. I did write some stuff down as well. Sorry, like I said, I was gonna prepare a little bit, so let me get that up as well. So I think when it comes to accountability, I think I told you about some of the main points that came up on like the principles sheets from some of my past workplaces and also like experience that I've seen since the real AI boom has happened, especially the recent kind of AI boom, it's really been brought forward in terms of the importance of accountability. For example, I was just reading last night about Microsoft and how their Copilot plus was maybe going to take screenshots every like 5 seconds of your- or every few few seconds, I don't know exactly, of your desktop in order to help you better, and so it had like context about what you might be asking it. But of course that has huge privacy concerns. And I guess accountability in scenarios like that would be transparency again, because if they didn't tell anyone that the Copilot worked that way, then it would be a huge problem. So transparency is a big part of accountability for sure, because if you hide the way your models work or what your meta parameters are and things like that, you can run into massive problems. So I think transparency is first and foremost there, and accountability is also discussing if you've made- also transparency when it comes to ethical considerations and doing the ethical considerations in the first place. So you know like, should I adhere to ethical guidelines? The system in general? I suppose there's multiple layers, like if you're using a packaged machine learning model that's been made by someone else. If you're not training it yourself, you should be forced to kind of look into the ethical considerations they've made because you can shoot yourself in the foot if you let someone else handle a certain level of it for you. And then if you're training it yourself, you've got all the traditional, like, ethical considerations you would make on any data set. And I think in most places I've worked, it's not only there, like accountability in terms of that, but it's accountability in terms of the data and of

course to use machine learning and AI, you have to use data. So ideally, there's already been a separate accountability and ethical kind of, how do you say, ethical kind of assessment of the data set or the data as well. And then separately, I would hope there'll be another kind of assessment of the AI system or machine learning system you're going to use in case it manipulates that data in a way that the original people didn't think was possible. In terms of regulation, adherence to regulation, I think we talked about that in some of the initial questions. It's really important. And then especially for banks and the financing industry in general, adherence to regulation is first and foremost. If the European Central Bank asks you to comply to a stricter set of standards than the rest of many industries, then you just have to do so. I think you can even be fined in a lot of cases by governing bodies such as The Netherlands Bank or the European Central Bank, if you don't adhere to their guidelines and audits, because we will often have questionnaires where they will come and they will assess us on how we're doing and if-

Interviewer
Yeah, like yearly, right?

Interviewee
Exactly. Yeah, I think so.

Interviewer
Yeah.

Interviewee
Accountability. I think you should be accountable that it's secure as well. The security should come in to accountability and robustness. Imagine that we start using them for like really- Robustness is like if we start using them for things that could go wrong like, I'm struggling to think of an example, but let's say we start using them in the police or in the medical or something like that and it's used to identify like cancer or something, and then the model breaks or becomes really untrustworthy. And then we can't identify cancer using the traditional system for like a month and people could die. You know, there should be massive accountability in that in certain sectors, and the finance sector comes into that as well. Like 'banks work on trust' is what people say a lot. It's like a common phrase in our bank and every bank in Europe, like that banks work on trust first and foremost. And I think that is really important, looking at these issues as well.
I think that's it for now, yeah.

Interviewer
How do you-
I think that's a nice resume of like how, what your view is within the use of AI systems, also in the financial sector. Transparency, regulation, ethics, robustness, I think was mostly the main terms.

Interviewee
Yeah, security and robustness kinda go together, yeah.

Interviewer
Yeah. And how do you like experience accountability within your firm? Or sector.

Interviewee
So in terms of if I see- I see it positively for sure. And then I see it controlled through a mix of trainings, rules, and then questionnaires, and audits, and brainstorming sessions. Well, brainstorming sessions maybe happen at a different level. And then there's separate, what comes last I guess, is assessments. And assessments may come from those inside of an organisation that you work for, that have the capability or who have been deemed to make the final calls on whether something that uses machine learning or AI will go to production. And then maybe they also make a call, whether it's used in certain contexts or if certain risks that you've identified or if you've missed any risks as well. If they see risks that you didn't already see, that can be really problematic actually, because then you don't have a solution ready to tell them. And that can be a good thing, if it's something you didn't intend and you don't want to rush out with like, haphazardly trying to cover risks, especially when it comes to machining and AI. Yeah. So I would say that's how I see accountability in adherence to all of the things I mentioned; being as open as possible and stuff.

Interviewer
Yeah, and like the questionnaires, is it like a some sort of way to measure it as well, for a person to assess and then the accountability is measured in a sense or? Are there other ways to measure it?

Interviewee
I would imagine and from- I would imagine that there's coverage. Yeah, that there is coverage percentages and then there's- if you've identified like, let's say 100 risks, you would usually have to have scenarios. And I forgot what it's called, but it's when, it's in

terms of resilience for example, like what your backup plan may be. If this or not- if this part of it couldn't work -I don't know, let's say it was a chatbot or something. What would you do if the chatbot did not get approved due to it potentially hallucinating an answer and the bank says that's not okay. What's your back up plan? Will the application still go to production? Do you have a more concrete chatbot that has like- that does not use like a large language model and has more of a local machine learning base like NLP, transformer model, or something that is trained on the bank's data. Instead of using like a RAG (Retrieval Augmented Generation), for example, which might lead to a lot of problems. So let's say you did want to use RAG. Maybe you suddenly find that it has like a hundred more risks, some of them high risks, some of them medium risk. You've listed out on all of those what you think the risks are, who the risk might affect, what the risk is. And in terms of risk management, you fill out like several other deals like how you cover the risk, if you accept the risk, or what you've done to fix the risk, or to make sure the risk can't occur. And then that's all part of the risk assessment. And then, yeah, then you have your entire solution design and then people from different disciplines comment on the solution design to see what the problem and see well see what other issues might be there. So yeah.

Interviewer
Hmm. OK. Yeah. So there's a very clear assessment of all the possible risks that could occur, and how they can be mitigated, and that's also reviewed?

Interviewee
I think so, but if you imagine that someone, some team controls that process and it's ever changing and getting added to.

Interviewer
Yeah, yeah.

Interviewee
I assume that there's some questions there, in these security questionnaires and the questionnaires you get when you're initially onboarding an application, that definitely were not there two years ago, before the rise of LLM’s in the last few years. Yeah.

Interviewer
Yeah. So it's it's been evolving quite quickly.

Interviewee
Yeah.

Interviewer
Perhaps it's unclear if the systems that were implemented five years ago, if they also adhere to the new set out regulations?

Interviewee
So I have to be careful here, but in most applications in the finance industry, they will get given- kind of- It will get given- In security, there's something called the CIA triangle: Confidentiality, Integrity and Availability. And you can have a score like 111 or like 112, and that is how a lot of like the industry talks about the risk of a certain application. And you can look it up and like maybe like a 111 application is like; the availability must be available because it serves transactions. So let's say that's there, and then the confidentiality, it's people's financial transactions, so maybe that's a 1 as well. And then integrity as well. The data needs to be completely- it can't be tampered with. Maybe it has to be on a system such as mainframe or on premise, and we don't want that data to be on like the cloud somewhere. I'm not saying that's the way my current organisation handles like each of those scenarios, but that that can definitely be a deciding factor on, to answer your question, how frequently an application is re-evaluated. So with the 111 applications, I can't tell you like how often, but maybe they are evaluated again against the newest set of standards quite frequently. Yeah. So they have to- and so then it becomes, you know, that team has to answer questionnaires and the assessments. There's these things called the CRA's; Change Risk Assessments. And whenever they make a change to the application, they have to log it and they might have to do another risk assessment and that is a common thing in the the finance industry as well.

Interviewer
Okay. Interesting.
And like in terms of implementations in your view, what should be run by AI, like what solutions or systems or processes, and what not?

Interviewee
Okay. I think Copilots- we're in the age of Copilots, as the GitHub CEO currently said. We're not in the age of, like generative AI, as in AI thinking for itself. And he's trying to quell the recent kind of freak outs about people being terrified of AI, AI taking over and things like that, and I do agree with him that the Copilots are getting pretty good. So, for like Microsoft's main example of in day-to-day work there's copilots that can assist

people's productivity. I think improving productivity inside of businesses is one of the biggest and the most profitable use cases that everyone is rushing towards. The summarization of like phone calls automatically from transcriptions. The automatic summarising of documents, like for example your company writes the end-of-year-report every year in a similar way. You just give it a folder full of images, you give it a few infographics, you give it a few graphs, you give it the financial reports and you say 'write this exactly like I wrote last year's report' and it'll just do it. And that could have saved several weeks of work of someone who is a chief financial officer or his aides who are getting paid in high compensation. And because of how much we pay humans, the time saving factor in terms of efficiency and productivity for these models is worth a lot of money, because they're taking the jobs away from the humans. So yeah, that is definitely where it should be used and where it is being used. And in tasks that kind of fall into the 'can be done by humans, but don't take huge amounts of understanding'. Cause at the end of the day, they are just transformers and textual N-grams, and like similar E cosines and matrixes, trying to come up with predicting the tokens that are going to come next, kind of thing. And when it comes to like 'oh, I want to teach ChatGPT about this'. That's the kind of thing we have to be scared of. Because people who say things like that, don't actually understand how the models work and they don't understand what's going on behind the scenes.

Interviewer
Yeah, yeah, they see it more as a person then or like a movie type of robot, instead of it's really, just as you said, a bunch of graphs and formulas and predicting models.

Interviewee
Exactly. There's a rush towards it right now and that is what I think a lot of people are scared of, like the fast implementation without careful thought for accountability and proper onboarding and regulation and what companies are gonna be able to implement this into their scary, scary things, as I said before. So like I said, like medical, police, kind of like anything to do with allowing it to judge humans. I would say, that kind of thing we've got to be very wary of. Anything where it can make a- it can be put into a process that can affect the humans and the way that they're affected, whether it's through like it makes a medical assessment of you or it makes a psychological assessment of you or like, assesses if you're likely to commit crime or it assesses if you should or shouldn't be able to get a mortgage on a house, for instance. We use mathematical model models for all of that. But at what point do we start to worry that people who don't understand the models are implementing these in the wrong way? And that's why the governance and regulation has to be really high. 'Cause yeah.

Interviewer
Yeah. Yeah. So that there's like still a human watchful eye on it?

Interviewee
Yeah, because LLM's are trained on billions of- tokens is the wrong word, but like- Because they're trained on such a huge data set and that data set is closed, you don't know exactly- you, you, we start to- people start to take shortcuts because they didn't bother labelling and training on datasets that they can see. If you have a data set on e-mail and you say like 'these emails have good context' and 'these emails have bad context' in terms of like human behaviour people are angry in these emails, people are happy in these emails. If there's a label data set and there's column that says good and bad then you can start to do data science and you can start to train. But if you've got an LLM and you just hand it a million emails and you say: please quantify these in terms of- classify them, sorry- in terms of like good and bad context, like human context inside the emails, then that kind of thing's worrying. And let's say it just happens to get a good result, because let's say it didn't even write code to do it. It's just doing it from its knowledge. Who knows what it's hallucinating. Maybe people have talked about datasets on the internet and it's read the reports. Maybe it's literally scanned, like Hugging Face and scanned some of the online data set repositories where people talk about data sets. Maybe it's seen the data online and that's why it's doing so well at classifying them. You have no idea, and maybe it's referencing something completely wrong and you trust it to make assessments of people but it's not a real model, it's someone else's model on someone else's computer that you're sending an API request to and you think it's doing one thing, but really that's not how it works at all. That's something that I'm really scared of, like when it comes to how fast we're advancing right now and-

Interviewer
Yeah.

Interviewee
'AI as a service' is what it's called, where you get an API key and you send it out.

Interviewer
AI as a service? Yeah. And it's basically, I think how you describe it, I think the mystery behind it is, what is scary about it, like that a lot of people do not have a clue what it entails, and even the people that do have the knowledge, are not sure what it is based on what the machine decides basically.

Interviewee
Yeah, you should know what it is, yeah.
You should use it for things like like we talked about before, like if I wanted to make a regex for a timestamp. And I could Google like the timestamp and be like give me the- 'Does anyone know the regex for this timestamp?' Maybe someone's asked that question on the Internet before, maybe it's been discussed somewhere on a forum, even on a coding forum, or on a tech forum, or on some regex forum. I don't know. It's been- it's on the Internet, all these regexes. And maybe there's even been discussions online like, 'oh, that one doesn't work for these reasons, it should be done like this'. So you know, at least when it comes to large language models, if I say like 'give me 100 regexes for all time different timestamps and try and make them all usable in like regex for Python', it'll do a pretty good job of like taking notes and taking references online about how regex for Python works and then it will find the other regexes as a separate task and then it'll, you know, spit out all of these regexes for me. That kind of thing like massively saves my efficiency and increases my productivity. And those are the kind of tasks I see it being like really used for. But like making assessments, you don't know where the data is coming from, so how can you like, really trust it? Yeah.

Interviewer
Yeah, if you don't have direct proof that it's 100% accurate, it becomes hard. Yeah, how do you trust what comes out of it, basically.

Interviewee
Exactly, you know, 'cause I could say I could be building a system that decides whether or not people get loans. That is not what I do, by the way, this is complete speculation. But I could say what features might be important if I have a data set of people's like recent financial history, their credit rating, and all of these other things about them, and imagine where they live is also in there. And then, yeah, maybe it spits out a bunch of features for me to use. Like if the average bank balance has always been like minus 500, it's going to say- it's going to make a correlation if I give it a label data. But then let's say it just gives me the code for that feature. Then it's up to me. Like I'm not giving it this person's data. Let's say all the data I gave it was like obfuscated data, like completely fake data. But it gave me this feature and I can decide whether I use that feature in my programme that I'm writing separately and it's up to me as a developer, because that's what I do, to ethically use that code. If I close my eyes and blindly copy and paste it into my code, that's no better than me just doing the exact same from GitHub. The code might be from GitHub as well, that that model, that an AI has just given me back. So yeah, those are also really bad scenarios; people closing their eyes, blindly copying features that had been suggested, or it doesn't think about- you didn't tell the model to

think about it ethically, or the LLM to think about it ethically, the ChatGPT for example. And then let's say it gives you a feature that says ‘these people of this post code no, don't give them loans’. Like because of the data you've given me and that’s just ethically wrong. Completely, yeah.

Interviewer
Yeah, it's the same as the taks fraud thing here,

Interviewee
Yeah, exactly. That's why I took the example.

Interviewer
But yeah, I think you mentioned it before as well like that in [Company], they are also trying to do like the risk assessments that the employee does themself, I think. And is there any like further active role that the organisation is taking to implement accountability features in AI in that sense?

Interviewee
Not allowed to talk about [Company] in that kind of level of detail, unfortunately. That's the only one of the only questions where I can't talk about what specifically they do and do not have right now.
I think I've talked about like efforts and stuff that might be going on, like may or may not be and stuff I've heard, like hearsay and kind of like assumptions. And I have confirmed some of it for you that I've seen it happen in industry and like this kind of things are happening. But in terms of like exactly [Company]’s process when it comes to AI accountability, I can't say unfortunately. I do have to kind of pass on that, but it's very similar to the- It's very similar to the ones we've talked about and what I can say is that with the increase that we talked about as well, in the use of LLM’s, that they are adding more and more to their to their process that you would have to go through when you make a change on a high risk application or if you want to use AI and things like that.
And you can see the public information about like, we’ll work with Microsoft and OpenAI and how we've been using that, I think I mentioned last time as well, and on how we've been using OpenAI to summarise conversations.

Interviewer
Yeah.

Interviewee
And that is a really interesting one on how they've had to manage that because essentially, at the start of the phoning conversation, you're already giving consent for the conversation to be recorded for training purposes- for training and some other purposes. I'm not exactly sure what the prompt says at the start of the phone call, but yeah, if the summarization of phone calls can already be done from the notes that the human takes on that, or notes that are automatically taken, then yeah, we already do have some examples of AI being used in production there and it's going quite well. And we work alongside Microsoft on that, and they are responsible as its AI as a service. They're responsible for a lot of it. I don't if you know but, you'll be able to find this online, but Microsoft recently said that if you get sued for copyright because of the output of OpenAI as a service, it's on them completely. It's on them. Yeah, yeah.

Interviewer
It's on them, you said? Okay, yeah.

Interviewee
So if you generate an image or something with Dall-E or one of their official models or computer vision, and someone tried to sue you, it would be them versus Microsoft. It wouldn't be them versus you. And that is one of the protections of having OpenAI as a service as well, that you are protected because the models are on them. Such as like the New York Times, like suing them or something. You're not going to get sued. Microsoft will get sued. So you know, if you did want to use you own GPT kind of model and you did train it yourself. Maybe you would regret not having some of the protections if you were a small to medium sized enterprise, as what you get when you use AI as a service.

Interviewer
Yeah.

Interviewee
Yeah.

Interviewer
That's very interesting, though. That Microsoft is putting their money where their mouth is.

Interviewee
Yeah, and who knows? Maybe they'll take it back one day, if they start getting hundreds of lawsuits, thousands of lawsuits. But but right now that's the case. And they also, you know, they reiterate 5, 10 times in every meeting that your data that you enter into their large language models won't be used for training. Ever. Yeah, that's very.

Interviewer
Yeah. So that's nice, that, basically, with using Microsoft's AI as a service, you also get a package of governance with it and-

Interviewee
Yeah.

Interviewer
They try to, in every way, make it as ethical use as possible and also put like a force behind it saying ‘okay if it goes wrong, it's on us, it's not on you’. I think that also creates a lot of trust for companies to use it.

Interviewee
Yeah. Exactly. Obviously it's not a real problem for the finance industry, but if I worked in the AI industry and you launched a model and it was used for like very bad things like that's also a huge problem. So like censorship and stuff is all handled by Microsoft as well. And like what people can and can't ask these models in terms of like, I don't know, I can't think of some examples right now, but bad things. Yeah, that's all controlled there. Like for example some- one of the biggest fears is the use of deep fakes by AI and that is a huge concern, right? I've already told you about the threats to our industry last time about how the people can make phishing emails and yeah, there's also the same concerns with people making deep fakes that can, you know, show like fake assassinations, fake news, fake stock information and calls like financial imbalance. And there's a lot of problems with deep fakes that everyone is concerned about. So it depends if you're the main provider of AI as a service and you can handle the censorship of those things, everyone else- and you reach the global standards and regulations, maybe it's easier to side with AI as a service instead of building your own complex AI models to use, which you then have to prove regulation for. Yeah.

Interviewer
Yeah, yeah, true.

And. I think it touched on it a bit before, but what is your overall view of the success of using AI for different processes?

Interviewee
Yeah, as I said before, efficiency, productivity massively increase and you've got saving costs massively increased and then of course I think that's just about LLM's, so I suppose it's a bit broader than that, but the the applications of how we can apply it in general in computing are going to be like monumental I think. Then the amount of data we can train on is growing. Computing power is kind of like levelling off a bit, you know, ever since we got to like 4 nanometer, 5 nanometer transistors. If if the will keeps going the way it's going and, you know, there's no war in Taiwan and like transistors will stay being created, keep being created at the rate they are and we keep producing like graphics cards at the rate we are, there's going to be more and more AI. I don't know what it's going to do for the environment. Potential bad things there, you know, green computing is something we're very concerned about and so is greenwashing. I don't know if you've heard this statement, saying that you're like a green company when you're not is very, very important to the companies I've worked for, let's say. So any statement you need to say about being green is really important. So if you take a load of- if you save a lot of cost, but you start killing the environment, personally in my own opinion, I don't really agree with that. Like if you like quadruple your energy spending to save costs in other areas, I would see that as negative. But then in terms of AI usage in physics and health, you know huge positives to gain there. And those are really promising and really exciting, for me especially as a big nerd. But yeah, I think that's about it. Yeah. And the things we've talked about before, but I don't want to reiterate too much.

Interviewer
Yeah. So I think there's a lot of positive things and a lot of change, to summarise it, but also risk as well if people don't use it or don't research enough what comes out of certain systems and if they can fully trust that, or if they need to do some research themselves on if it's right, basically.

Interviewee
Yeah. It kind of- it's saving you hard work on so many levels, right? There can be like the personal level, the employee level, the small medium enterprise level and large organisations making mistakes as we've talked about of like Air Canada. And then Microsoft, even at the top of the food chain recently being rebuked about the privacy concerns about its Copilot plus feature that takes screenshots all the time. That news just came out and they are supposed to be like the leader and they have one of the

strongest partnerships with OpenAI, who is the market leader right now in AI, and possibly future generative AI. So yeah, it's at all levels of the food chain, this complacency and adherence to privacy, regulation and governance, and there's issues all over, and I don't think it's the end of it either. I think this year I wouldn't be surprised if the amount of news stories about people messing it up through the roof.

Interviewer
Yeah. And then can you? Have you experienced any incidents that you can mention or?

Interviewee
No, sadly I can't.

Interviewer
No? Okay, that's no problem. And then there's a, so yeah, there's a more general question, but should people that use AI in their processes in an organisation be held accountable for what the AI does or decides in your opinion?

Interviewee
So this is in my opinion. And I think they should be held accountable, depending on their risk assessment and what they wrote. If they said there's no risk, then they should be definitely held accountable. If they properly presented these risks without obfuscating or like making any broad- like making it too broad or kind of trying to trying to hide what could happen, and it was accepted by the organisation, I definitely don't think the employee should be should be accountable. I think the organisation then and its process and its regulation and governance adherence is the problem. And then, yeah, if the organisation didn't have any of that, people shouldn't be easily able to skip things, they shouldn't be able to just fill out- if the team and the employee filled out even a form and lied on the form then they are accountable, I would think. If if an organisation had like a process to check it and one of those parts of the process initially would be ‘does your product use machine learning or AI?’ And they tick no, and it did, they're accountable. Because the rest of the process to do with governance of AI and ML might not have been followed because they said no.

Interviewer
Yeah.

Interviewee
And we will, I think, scan repositories, scan code for usage of these API’s and machine

learning, and any kind of graphics cards or CUDA (Computer Unified Device Architecture) stuff that's going on in organisations, that is unauthorised soon. I think I could easily see that happening. Personally I have no idea if it's happening at my organisation, it's not my department. It it could be, I can't say, speculating. But I would say that I'd be surprised if that kind of stuff doesn't get shut down, unauthorised use because it's like I said, if it's using it, it's a certain like CIA rating and it should be getting redone every time the regulation changes or every X amount of time there. Yeah. Pretty crazy stuff, but I I would think that, like we talked about the scenario, if I copied code made by an AI and it was full of errors and then I implemented that code separately and the AI helped me write it. Then I'm accountable, not the AI. I can't be like, 'Oh well, Copilot wrote me that code'. No, that won't stand at all because AI might not be used in my product, my end application, but AI might have been approved for Copilot. And I can tell you right now, in the GitHub Copilot at any organisation that's running- this is speculative as well, not necessarily [Company], but there's probably a huge policy you have to read as an employee and you have to sign off that you won't use it on data that interacts with customers, most likely, maybe you have to sign off that you won't use it on production level code. Maybe you have to sign off that you double check and read it. All you have to sign off that it adheres to like our company's coding standards and there's loads of things that you have to sign as an employee, that you will do if you use Copilot. They try and make it so you can't just blindly use it to smash together code that will end up being used on production financial systems, yeah.

Interviewer
Yeah. And that's also, I think you also mentioned that in our first conversation, that it actually is the only way that they can control the use of it, because if they wouldn't implement the Copilot, then perhaps it would be used unethically or-

Interviewee
Yeah.

Interviewer
Yeah.

Interviewee
Exactly. Covering- You gotta cover your own back, but also not prompt people to illegally use it and start increasing the chance of- Information protection, I think, is how we phrase it. Which is where people start to e-mail themselves code and take code on USB sticks, so they can put it in their personal ChatGPT subscriptions. That's a definite- that was not a risk, well in general it was a risk for information leakage like a while back,

people like sending stuff to their home computer to maybe, I don't know, use their favourite code editor that was banned at work. But not many people were doing that. But I can say now from what I've read online as well, across the industry, people are sending so much data to personal devices so they can put it in a ChatGBT and that is such a huge risk for the industry, a massive risk. And the amount of people doing that that I've seen online and in reports online are massive. Yeah, huge, huge risk.

Interviewer
Yeah. Is there a way to mitigate that you reckon or not?

Interviewee
Yeah, I think so, yeah. Make regulated Copilots in your organisation, and forced trainings to anyone who has to use them about the dangers of information leakage and information protection, yeah.

Interviewer
Yeah, so governance and knowledge-spreading, basically, yeah, yeah.

Interviewee
Exactly.

Interviewer
Yeah. And what are your thoughts on the ethical implications of AI driven decision making in the financial sector?

Interviewee
I think I covered this a bit when I was talking about like people getting mortgages and stuff, so I don't want to reiterate it too much. I don't think they should be involved in decisions that can affect clients too much.

Interviewer
Yeah, that could become discriminatory, or yeah.

Interviewee
Yeah, exactly. That's a good way to put it, indeed. Like any kind of discrimination and things like that, I feel like should always be done- It should be done by a model and it should come out with, like calculated figures. And then I believe that an analyst or set

thresholds should decide, based on carefully evaluated mathematical models that are made for the purpose of taking someone's financial history and giving them certain scores on features. And not made in a way where it's- maybe it's not being trained or trained on any discriminatory features. Because you know, you don't know what people's circumstances may be. But I cannot say like how they, how those decisions are made in my bank. And if they involve any kind of AI or ML, because I honestly don't know. But yeah, I would say most of the financial industry is using machine learning for decision making. And to what point new LLM's are used, I don't think very much across the whole industry at all, so.

Interviewer
Yeah, so ethically seen, they're being very careful, probably, from your point of view?

Interviewee
I think so. When it comes to large language models, but dedicated machine learning models that have been created to let's say, to put classification and anomaly detection and things like that and fraud, fraud detection is very difficult as well. Detecting fraud, most likely in most organisations, if they're an advanced bank, I would say they're using machine learning and fraud detection. You know, and that's because it's very hard to catch in financial transactions. We don't have humans trailing through financial transactions, as you can imagine. If you start to send strange patterns of money around, I can't say what, but you can be sure that these patterns will be detected and that's the essence of anti-fraud and financial crime. And it's definitely used in those kind of decisions to beat the criminals and to aid us in our decision making there. But highly unlikely that those models have been built for that as well, and the risks have been assessed and you can be sure that if any like action was taken against anyone, the ML part would have just been for the initial detection and that would have most likely sparked off a human driven investigation process. Do you know what I mean? It's not like 'An ML model has detected you, you're going to jail now'.

Interviewer
Yeah. And that's perhaps also the way they address accountability from that perspective, right?

Interviewee
Yeah, same in cybersecurity. You know, maybe the detection was built on machine learning, but the- it's we're moving towards, I think in in 5 to 10 years, where responses, automated responses might happen from machine learning. But if there's risk in those responses, I highly think- I'm highly inclined to think that a human would be involved.

Yeah. Like, you know, if the decision was to turn off something in a bank, for example, like this is all speculative again, we don't have this. If ML could be given the decision to like quarantine a part of the bank and it could result in people not being able to make card transactions, can you imagine how bad that would be? What if it got set off by accident? What if there was a three day weekend and that Kings Day was moved or something, and it was like, 'oh, something's going wrong in this application, better turn it off'. So humans should have the last call. It should be-

Interviewer
Yeah.

Interviewee
It should be mostly in detections and then you know, these processes, maybe ML's is just used in a small part and that increases productivity and efficiency, because imagine if I had to travel through billions of logs every day. That would be awful. Yeah.

Interviewer
Yeah. Yeah. And where do you see, like in terms of the future, where do you see AI being involved with your work?

Interviewee
I think I'll use it everyday in code assistance. I think that it'll be used in security sections in future and responses, as I said. And I think it'll be used for, in my work, automated normalisation, passing of logs, security incident tagging, information and event management, classifying how damaging an event might be. And yeah, we're in the age of data, it's like increasing massively and automatic parsing of the data and search is really important and I think yeah, I think machine learning and AI will help me in that as well.

Interviewer
Do you reckon your day-to-day work would change a lot then? If it would take over those tasks.

Interviewee
No, I'm pretty- How can I say this? I'm pretty high up the engineering chain, that my tasks are not repetitive at all. And I code different complex problems consistently that I find can't be solved by GPT's 'cause, as I said, I'm allowed to use them within my organisation as I'm part of a pilot and I've carefully adhered to using them in certain

scenarios, where it's safe to do so. I can't go into the exact policy, but that's that's kind of the gist and and I can already tell that at the moment it can't do what I have to do. It can help me though, it can speed me up. Let's say you make a recipe for like baking a cake and you need to put something in the oven twice and you already wrote the putting it in the oven part of the script. It can help you, you know, make that very modular. It can parameterize it so you can put something else in the oven, and they can do that in like a second. So you know, then it just saved me 10-15 minutes of coding. Which is super nice. And I welcome that absolutely. But will I let it write me a feature that I don't understand? No, I will not at all let it do that part for me. I won't even let it write a regex for me without testing it a million times in a syntax checker. So yeah, I see it helping me massively, but in my in my row, it can only assist me. Yeah.

Interviewer
Yeah, yeah. And as a final question. Should humans always be held accountable for the actions of AI?

Interviewee
I think so, yeah.

Interviewer
Also in the future?

Interviewee
Yeah, as I said. Humans, yes. Employees no. Well, unless in certain circumstances, sorry. Because, yeah, it's up to us, I think. I think they'll be used against us at some point. It's talked a lot about in cybersecurity right now, but AI's, and you know, we saw worms in the past in the initial days of cyber security and we talked about them as if they could change and self replicate. But that wasn't so true. You know, we train, we would code multiple ways for them to get around computer systems and spread from computer to computer. But if we get to the point where it writes better code than us, and we try and write patches for IT systems and it can outthink us and detect what we've changed and then create a completely new vulnerability that we've never heard of in an IT system, and it can just spread to every computer on Earth, you know. That's- even though I'm an IT expert, and that sounds very 'Terminator'-esque, it's a possibility that it might- that we might be outdone by generative AI, and if we don't prepare enough through governance and this project XPT thing that Google has I think, that Google and a few others are involved in, and we don't create these like huge regulatory bodies- Like there's a reason the governments have started to create these think tanks about how we can limit, and they use the word limit, like how we can limit the research if we need to.

Because it's really scary, yeah. And if one nation state can use any technology, whether it's machine learning, or AI, or something else, like some massive physics breakthrough or something, to potentially like have first strike capability on everyone else, it's going to affect us all massively, so.

Interviewer
Yeah. Is it even possible to limit the evolution in that sense?

Interviewee
Yeah.

Interviewer
I don't think there is.

Interviewee
No. If there's- You know if we get to like fusion energy, which solves our energy crisis, which is rapidly getting made worse by- which could be made worse by AI and machine learning and the amount of like power we have to- because you know, as we're reaching a point where the amount of power we put into something to make it like produce more computing power, it's kind of been a graph for a while, and now we've reached a point where we have to put a lot more power in, to like, improve the computing bias it in more percentage. And it was quite relative for a long time, and now we've reached a point where we're struggling to make transistors smaller, and we have to put a lot more power in, to improve the computing power. And yeah, if we have fusion energy and if one country gets there first and takes off in terms of science and technology, ahead of the rest of us, it could have, like, dramatic consequences for us. So yeah, in terms of if I think it's possible to limit it, no, I don't think it is because it’s hard to limit the research of something like that.

Interviewer
Yeah. Yeah, it's basically like the moon landing or something. People just want to be there.

Interviewee
We promised to- Yeah, we promised to stop testing nuclear weapons, yet you know, I'm not gonna get into that, but there's still nuclear weapon tests by some nations. So how - Are we gonna stop? Unless we all agree, then I don't know how it's going to happen. Yeah.

Interviewer
Yeah, yeah, yeah. No, it's. A bit morbid, no? Ha.

Interviewee
Sorry, been a bit dark and crazy there in the interview at the end, but that is like- that's I put on my my tinfoil hat there at the end of the interview.

Interviewer
Yeah, no, but I completely agree with you, 'cause, yeah, it's basically like such- I think the comparison with nuclear weapons is very agreeable, because how can you limit something that there's, like, a running competition for? You cannot, and it's basically only going under covers, probably, and even harder to regulate and see what's going on, if it's done secretively.

Interviewee
Exactly. And you can guarantee that it will be done behind closed doors, partially at least not only in academia yet, but also in private. This kind of research. Yes, I'll be very interested.

Interviewer
Yeah. Yeah. Well, we'll see how far my framework takes takes us. Maybe it will solve everything.

Interviewee
Nice. OK.

Interviewer
Well, thank you so much. This was it. We did it.

Interviewee
Yes, it took much longer than expected, but I hope all the extra talk is helpful. Yeah. OK, that's great. That's great.

Interviewer
Oh yes, it is. It definitely is. Yeah, yeah.

Interviewee
Alright. Yeah.

Interviewer
Yeah, it's way better than 15 minutes of somebody just saying yes or no. So no, it's it definitely helped me out a lot. Thank you so much.

Interviewee
Nice. No problem. I'm glad. I'm gonna go. Enjoy the sun now. I hope you can do the same.

Interviewer
Yes. Yeah, I think I am gonna for a little bit, yeah. Thank you. You too.

Interviewee
Nice. Alright. Have a good day and a good evening and take care. I'll speak to you on the other side.

Interviewer
Yes, they are to rammy for me.

Interviewee
Right. I will see you in a bit.

Interviewer
Bye bye.