Opening Statement

Interviewer:
Good morning and welcome to my interview. First of all, I would like to thank you for being willing to participate in this interview. My name is Imane Akhyar and I am a master's student in Information Studies at the University of Amsterdam. This interview is part of my thesis research, which I am conducting at [Company Name], on how artificial intelligence can be effectively integrated into cybersecurity risk management within financial institutions. The interview will last approximately 60 minutes and will be recorded for research purposes. Your answers will be completely anonymized and kept confidential. You may skip a question or end the interview at any time. Do you have any questions before we begin?

Respondent:
No, no questions at this time.

Section 1: Background of the interviewee

Interviewer:

Q1.1: Well first introduction question, can you briefly tell us a little bit about yourself and your role within the organization where you work?

Respondent:
Yes, that's fine. [Name], I work at [Company name], where you are also writing your thesis. Obviously a firm in governance, risk and compliance, mainly within the financial sector, but also outside of that. Although I work mainly within the financial sector, even more specifically mainly with pension funds, but also other clients. We focus mainly on the IT side of things, but with certain clients we also do a broader approach, i.e. enterprise risk management. I myself studied at the UvA, the same course as you, Information Systems as a master. Then I started working full-time at [Name of Company] almost two years ago now. So from the direction of management work and risk analysis to DORA implementation: how to reflect that in policy, how to reflect that in measures. Pretty broad. And lastly, what will it be, the last few months, maybe four, five months, I've also been busy with AI: the opportunities and threats within our company, and also a bit of what we can do for that with clients

Interviewer:
Okay, great. So you had mentioned that you've been working there for about two years.
Q1.2: **Is it also that many years of experience that you have with cybersecurity, risk management or AI? Or do you have even longer experience in those areas, for example?**

Respondent:
Well, well, that all kind of goes back to college. So I did the Master's in Information Systems and before that bachelor Information Science, that's where those topics come up as well. But in terms of work experience, it's really the two years at [Name of Company]. Something more specifically, I also did the CISA course, the Certified Information Systems Auditor course. I completed the exam of that as well, which does still give relevant knowledge about at least the audit piece (the relevance where AI is also going to find a place in it). But in terms of work experience: yes, almost two years.

Interviewer:
Okay, great.

Q1.3: And then you had also mentioned the DORA and the AI Act, so you're actually quite familiar with that then, the DORA and AI Act?

Respondent:
Yes, the DORA I know practically by heart unfortunately, and the AI Act is a bit newer and a bit more vague. But we're just busy working on that now as well, so I at least have a good idea of what aspects it covers and what the focal points should be. And in addition to the AI Act, for pension funds specifically, you also have the Code of Conduct on AI and Ethics for the Pension Sector, which is already trying to give a little more substance to the AI Act. What's interesting about that is that the principles mentioned in it are, as far as I'm concerned, more widely applicable than just for pensions.

Section 2: Thematic depth

Interviewer:
That's it then for the introduction. Now let's go to the core questions of the framework. It's broken down into certain components; first it's mainly about the elements themselves and the second about ethics and governance, for example, but we'll see that in a moment. The first question is actually pretty straight-to-the-point, goes right all the way through the framework:

Q2.1: At what points within cybersecurity risk management or risk management in general do you see the most potential for LLMs to add value?

Respondent:
Yes, it's mostly a very open question: so at the point of cybersecurity and IT risk management: where are the opportunities for AI? Whew. Well, well. There are a lot of opportunities, but that's kind of the problem: there are a lot of opportunities. So I'm just sitting down now and looking: what are the focal points? Look, obviously what you can do anyway is IT risk management actually starts from your own strategy. So what do you want to achieve as a company? And AI by itself can help with how you want to deploy AI, so to speak. It can already give you an idea of what the best practices are in the market in terms of technology use. Then you could reload your own strategy document however that comes back into your organization and ask: gee, what am I maybe missing? And in the process get the kind of tips for: how can you integrate this into your strategy? Risk management has so many parts. You can also ask for inspiration based on your policy documents and a description of the context of your organization: what risks might we have? And then again, how might we mitigate them? So there are a lot of opportunities. I do think that with all those opportunities, it's very important that you have the subject matter expertise to properly

assess. Because, at least at the moment, the information is not always correct. In fact, there are still a lot of hallucinations.Yes, I think this is a difficult question to answer "where can it help?" because I think: almost everywhere.

Interviewer:
Q2.1a: Okay yes, and then when I talk about LLMs, **are there certain types of LLM that you think are most appropriate within your own organization?** Or that you think 'okay, we can use these anyway'. Or are there others that you think 'they're a doubt'?
Respondent:
And what do you mean by certain types of LLM, say the standard language model, or a reasoning model, or?

Interviewer:
For example, GPT-4 and Claude, those kinds of tools .

Respondent:
Okay. Well, we have just actually determined which tools we want to use in the short term. For that, we made the trade-off of using ChatGPT and Claude as standard tools that just may be used, with the context that we are not making a paid version available at this time. For example, what is another opportunity of a Copilot, is that we have the Microsoft environment, we have SharePoint, we have Teams and that integrates super nicely with that environment. So yeah, that would be an opportunity: that you could very easily, if you deploy Copilot, say 'well, all that data is loaded and it becomes super easy for colleagues to ask certain questions.' Do we already have something for this? Can we create a template based on what we have? What's in our manuals?'
At the same time, that also brings another risk, which is that that has access to all the data. And if I say it right, Copilot has basic access unless you turn that access off. So if at some point you don't have that set up properly and you have secret documents that suddenly everybody has access to through Copilot, then you have a big problem. And yes, those are the types of risks that we then look at when weighing up: what do we deploy? But yes, so right now we've landed on GPT-4, because that's just a very advanced model at the moment. But yes, that rat-race is still going on, so they might not win it. Claude we use for programming. And Apple Intelligence, given we use Mac products and it's already integrated, which is yes, part of that you can run locally and that part can be used as well. So right now the considerations for that have been mostly privacy concerns: how do we make it as difficult as possible to accidentally put data in there that you wouldn't want to put in there?

Interviewer:
FQ2.1a: And has there been any self-experimentation with those three you just listed , i.e. ChatGPT , Claude and Apple Intelligence, or is that coming in the future?

Respondent:

Experimented, yes definitely. So within [Company Name] we have a team within the organization that is already working on it. Or well, mainly working on: how can we facilitate that colleagues can use it well?
And well, we've been actively working for a while now to also deploy it in our own work and also to try out some use cases. Partly: what can we do right now? And partly towards the future. For example: if you have a risk analysis of a customer, I cannot do that yet, because we do not use customer data yet, but if you do have that, what can you do with it? How can you process that? How can you make your own life easier?

Interviewer:
Q8: Has there been any thought within your organization about creating your own LLM? Or is that something where you guys think 'oh, we might do that in the future'?

Respondent:
Anyway: to really fully or really well deploy an LLM, you have to be able to use customer data as well. Because as far as I'm concerned, the biggest use cases are in the data processing piece. Or well, besides the brainstorming and inspiration, because you can do that without customer data. Although in doing so, it's also nice if you can outline a specific situation (you can't do that right now either). But the data processing piece helps a lot. For example: if you say "we want to start looking at accountability reports," well, we have specific points that we pay attention to. And for that we have to search a whole document. If we can say "well," AI does that first step, so those points that we always pay attention to, for that it indicates: where does it say, and what does it say approximately, in a way that is verifiable. But especially also: what points do I not see coming back?', because then you know that those are your focus points and you can start giving advice mainly on those. That just helps tremendously that makes the work a lot more efficient. And then we and our experts can focus more on the content, instead of doing operationally tedious work, so to speak.

Interviewer:
Q2.2: Okay. And when LLM systems are applied , so it can be used by anyone in the organization, are there certain, yes, procedures whereby there is oversight of how the AI systems (i.e., how the LLMs) are used?
Well, some of it you can set up technically. So it's also possible to give your files certain properties that you say: well, you can't load this. Further yes, monitoring usage. Look, what we are going to do anyway (however we are going to deploy AI) is regularly ask our colleagues how it is being used now. And we have it a little easy in that, because we have a team of 20 people. Yes, we just have very short lines of communication and we see how people are using it; in a large organization that would be a lot harder. So yes, we actually have the luxury of already having more visibility into how it's being used. We haven't yet looked at whether we're going to deploy a paid version or whether we would then have visibility into the prompts and into what data is being used. I actually wouldn't know; interesting in itself. But yes, insight into the prompts would be a bit scary for me personally. Because yes, I don't need to see what my colleagues are searching for on their work computer on Google, I find that very invasive. I don't think you should go that far.

Interviewer:

Q2.4: You just mentioned privacy as well, so that's one of the concerns. Are there other risks that you see when deploying LLMs, for example bias, discrimination or misclassification, that an organization might run into anyway?

Yes, 100%. Now, of course, the world has. Yes, privacy is one, quality is another. And if you look at bias, yes, every LLM is trained on certain data, and also not trained on other data. In the information sector (I'll call it that) one problem is that data itself is also colored. Certain people with certain characteristics will be included more in data. You see that in healthcare, for example: that almost everything in the past has been tested on men and very little on women, so we just know a lot less about how women should be treated. Yes, that's true with all data. We don't have generalized data sets, so there are no LLMs trained on fully generalized data sets either. So bias is there. Well, that's always inherent in AI at all but certainly with the data available in the world. That's just there, and that's something you have to understand super well when you deploy AI. What's also kind of interesting: the AI Act already indicates that when you deploy AI within your organization, you have to make sure that the people who are using it are sufficiently AI literate.

Well, then the question is: what is "AI literate"? There is one definition for that, which is still a little broad. But it does indicate that you can use the systems safely, know how it works, and on the basis of which your output is created. That's basically what it comes down to. And yes, I think part of that is also that you understand, well, that bias is in there, and so that control has to remain there. It remains important. I think it's going to be very difficult, especially with those LLMs actually just grabbing all the data they can put their hands on, and most of the data is colored. So that bias is just going to remain, and that's going to take a very long time to reduce that. So for now, the most important thing is to be aware of that and also start actively looking at whether you need to take that into account in the output you get and what you do with it.

Interviewer:

Q2.4a: Yes, I also had a question about human control. There are obviously certain decisions that an AI can make, and some say "okay, that can be fully automated" and others don't. **Do you think then that human control is still needed with every output, or that certain decisions can be fully automated, so then it's actually no longer necessary for another person (just a human executive) to check it out? So if you use an LLM?**

Respondent:

Well, it varies by situation. Look, basically I think: our field is a field of expertise. We advise clients on how best to do things, and everything professional, as far as I'm concerned, should just still be checked by people, even with the four-eye principle. That's how we do it now. If I deliver an end product to a client (say, I'm doing a draft of a policy), I don't do it by myself; someone is watching. If AI helps with that, first of all, it's important that I look at it to see if I agree with what the AI is saying. And after that, someone is still watching, so that doesn't change. Look, if you're talking about things like chatbots, yes, inherently there's no more control over that, because that's just AI talking directly to an end user. So in itself, that can be automated. Even then, I think you have to have a fallback option so that there are still people involved. And yes, there are certain

risks. You see, for example, that a lot of chatbots that are already being deployed are just not that good, whereas they could be if you give them certain resources and let a GPT talk to the customer. That would go a long way. But that would also open the door again for such a GPT to say something that is not true at all, on behalf of the company. Yes, or that people are going to try to get that GPT to say something weird and that he's going to say really dangerous things and companies don't want that. So it becomes very limited. It can be automated, but that comes with a certain amount of risk. And you have to weigh that risk and see if you can accept it. Otherwise something else has to intervene.

Interviewer:
Q2.5: I was wondering: how do you actually handle the data you enter yourself into AI tools, such as LLMs? **Does that data have to meet certain conditions before you use it, for example to reduce bias? Or are there no really clear rules for that yet within your organization?**

Respondent:
Yes, for the time being, no. I do think it's an interesting question. Look, basically the way we look at it now is also a little bit more of: eventually you want to start making data available. There's a part in "what are you not allowed to start using" and a part "what are you allowed to start using." And the "what are you allowed to start using" part yes, in time that will probably become more so. Probably just in the terms and conditions (when we enter into the contract with the client) it will say that we are allowed to use certain classifications of data. I think it's going to be tricky to say that you actually have to pass a checklist before you put a document into AI, because it has to remain approachable enough that it promotes efficiency. If you have to go down a checklist for every document you want to put in there well, that makes it a lot harder. We work with mega-many files that we get from customers all the time. So also if you're going to say of every file that comes in "Well, we're going to check it beforehand to see if it complies" that's tricky. Especially with something like bias; that can't really be automated, you'd have to go past that manually.
So I'm afraid it's unrealistic for us to say "it has to meet certain bias conditions." But yes, whatever I say, I think you always keep that bias, and that's not necessarily AI-specific by the way. It's already the same in every organization now, that bias is in there in some way. It's generally not intentional. The best thing you can do is to be aware of it and ask yourself from time to time "hey, what is the bias we are working with now and shouldn't we do something about it?". Yeah, for AI specifically it kind of follows the organization itself, I think. And to really go down a checklist, I think is difficult.

Interviewer:
Q2.6: And when it comes to the use of LLMs: do you guys have any clear agreements or guidelines on how to work with them? For example about things like prompt filtering or model training, do colleagues know what can or cannot be done, or is that still a bit of a gray area within the organization?

Respondent:
Yes, it's still a bit far away for us, I think. We're still in the exploratory phase, where step one is actually to encourage colleagues to use AI at all. And so we did training in that and we continue to do that with "how best to do that." We just had the first training, and in a month we're going to do another more hands-on training, including use cases and more examples of "look, that's how you can do it too." We have no plans now to limit how prompts can be used, except that we did agree among ourselves that prompts should not be traceable to customer situations or specific customers.

Interviewer:
Q2.3: I had another question about regulation and governance. Regulations like the AI Act, GDPR and DORA obviously change over time as well. How do you guys deal with that as an organization, especially when you look at using LLMs in the future?

Respondent:
In most organizations, especially in the financial sector, there is a compliance officer who is responsible for keeping track of and applying laws and regulations. It's the same with us. If something changes in regulations, like with DORA or now with the AI Act, that person picks that up. There will not be a separate team for AI compliance in my expectation; it will just be incorporated into the existing compliance structure.

What happens then is actually quite standard: we start with a gap analysis. So: what exactly does the new regulation say, to what extent do we already comply with it, and what still needs to be adjusted in policy or processes? Once that is done, internal audits are used to check whether everything is actually being complied with. That structure is already in place; there is no need to reinvent it for AI. You essentially have ‘Run’ and ‘Change’. Run covers what the organisation routinely does: in that mode the process begins with Risk Identification & Assessment, which is scheduled annually.
Seen from a Change perspective, a trigger, such as a legislative amendment, forces us to launch a project to adapt how we work. You don’t wait for the regular risk analysis; you immediately ask, what do we need to do now to meet the new requirements?

Interviewer:
Q2.7 and Q2.8: Finally, a question about the framework itself: what would you need to have real confidence in an AI framework for cybersecurity risk management? Do you think the elements that are currently in it are sufficient, or are you still missing parts? And how do you think such a framework should ideally be tested or improved in practice?

Respondent:
What I would need to truly trust the framework is a clear and logical mapping of the whole chain from risk identification to risk treatment.

In the Risk Identification & Assessment section you already list some very specific examples of how AI is used, but in my view that covers only a small part of the story.

I still miss a step back: which characteristics of AI require extra attention during risk identification? Based on that analysis you then decide which risks are relevant for your organisation and what measures you must take to mitigate them.

Those measures, in turn, give concrete shape to the laws and regulations that were created to limit larger, more generic risks. In other words, there is an intermediate layer between risk identification and your legal obligations where you translate both statutory

requirements and organisation-specific risks into actual controls.

Certainly, statutes like the AI Act or DORA provide direction, but in the end it is your own risk profile that determines what you need. I would like that distinction to be more visible in the framework.

In addition, I would also recommend including something about the type of LLMs you deploy where. Not necessarily a list of specific tools, which are constantly changing anyway, but more at a high level. For example, where do you use a general-purpose model, where just a reasoning model, and in what context is a retrieval-based model appropriate? That kind of direction helps make the framework more concrete and future-proof.

Finally, I would add the education piece. If the framework makes clear that full connection between risk analysis, mitigation, governance, and deployment of the right model types, that would really build confidence for me.

Closing:
That brings us to the end of the interview. Thank you again for your time and sharing your insights, this is immensely valuable to my research. As previously stated, everything is completely anonymized and kept confidential. In the second week of June, you will receive another short Google Survey, in which I will present the final
AI framework and ask some additional validation questions. This will allow you to indicate whether the framework matches your practical experience and whether there are any points missing.

In addition, if you would like to receive a summary of the findings once the thesis is completed, please let me know and I will be happy to send it to you!