Opening Statement

Interviewer:
Mijn naam is Imane Akhyar en ik ben masterstudent Information Studies aan de Universiteit van Amsterdam. Dit interview maakt deel uit van mijn scriptieonderzoek, dat ik uitvoer bij Kouters van der Meer, over hoe kunstmatige intelligentie effectief kan worden geïntegreerd in het cybersecurity-risicomanagement binnen financiële instellingen. Het interview is gepland voor ongeveer 60 minuten maar zal ongever 30-45 minuten duren en wordt opgenomen voor onderzoeksdoeleinden. Uw antwoorden blijven anoniem en u mag op elk moment een vraag overslaan of het interview beëindigen. Heeft u nog vragen voordat we beginnen?

Respondent:
Nee wat mij betreft niet.

Sectie 1: Achtergrond van de geïnterviewde
***Interviewer*:**
Top dan kunnen we beginnen.

Q1.1: Kunt u zelf voorstellen en vertellen wat uw rol is binnen de organisatie waar u werkt?

Respondent:
Nou ik ben [NAAM], ik werk op dit moment formeel voor twee organisaties. Of eigenlijk misschien zelfs 3, maar ik ben officieel in dienst bij [NAAM BEDRIJF]. Dat is een partnerorganisatie van [NAAM BEDRIJF]. Denk aan het reilen en zeilen van de organisatie, maar ook aan besluiten over wat we met de applicatie [NAAM] gaan doen en of je die kent. Maar ook hoe we dat gaan uitrollen bij partners.
Daarnaast werk ik een dag in de week eigenlijk een soort van gedetacheerd vanuit [NAAM BEDRIJF] bij [NAAM BEDRIJF] om bij bestaande klanten te helpen. En ik werk ook nog een paar dagen per week bij mijn voorgaande werkgever en ben ik aan her afbouwen. Daar werk ik als VP Technology & Security en mijn rol is betrokken zijn bij eigenlijk de AI initiatieven die we daar hebben. Zowel van hoe kunnen we binnen de organisatie gebruiken? Wat zijn de security consequenties daarvan? Maar ook hoe kunnen we het in het product integreren? En wat zijn daar dan de aspecten die nu aandacht behoeven?

Interviewer:
Oke top dus wat ik hieruit begrijp is dat u dan wel ervaring heeft met artificial intelligence en weet wat het is en inhoudt.

Interviewer:
Oké top!

Q1.3: En in hoeverre bent u bekend met LLMs of andere AI-technologieën in uw werkcontext?

Respondent:
Ja nou ik ik, Ik weet hoe ze conceptueel werken. Ik ben betrokken bij de keuzes van in ieder geval bij [NAAM BEDRIJF], welke we gebruiken, welke tool we gebruiken. Ik ken de concepten als bias, LLM, machine learning, al dat soort aspecten. Ik heb er geen specifieke training in, dus alle nitty gritty details moeten mij niet vragen, maar conceptueel weet ik wat ze doen en hoe ze werken.

Sectie 2: Thematische verdieping

Interviewer:
Ja top dat waren dan de introductie vragen het volgende deel zijn dan de kernvragen dat uit ja dat te maken heeft met het frame op die ik gister heb opgestuurd. Ik weet niet of je die al door heeft genomen, maar meeste vragen zijn of een combinatie van de elementen of vragen van losse elementen en op het einde van dit onderdeel is een wedstrijden ook nog even kort bespreken.

Nou, de eerste vraag is best wel een lange vraag, heb ik ook al gemerkt bij de vorige interviews, dus ik moet het maar even zo stellen.

Q2.1: Op welke punten binnen het cybersecurity-risicomanagement ziet u de meeste potentie voor LLMs om toegevoegde waarde te bieden?

***Respondent*:**
Zo een lange vraag en een brede vraag.
Even nadenken.
Ik denk dat LLMs op meerdere punten kunnen helpen, maar één is gewoon in de formulering van welke beheersingsmaatregelen wil je nemen en hoe sluit je die enerzijds aan op de praktijk en zijn ze ook testbaar. Een tweede aspect is, denk ik dat het dat LLMs heel goed kunnen helpen bij het begrijpelijk maken voor mensen die niet heel veel kennis hebben van cybersecurity en of risicobeheer. Dat ze dat op een toegankelijke manier kunnen uitleggen. Zonder dat je daar heel veel moeite voor

moet doen. Je ziet namelijk vaak dat vakmensen al snel in termen praten die voor niet-specialisten lastig te volgen zijn. Ik denk ook dat LLM is uiteindelijk op basis van configuratie of evidence veel sneller en objectiever is dan mensen kunnen beoordelen of iets in de basis veilig is geconfigureerd. Ja, dan te nee en of maatregelen effectief zijn ja, dan te nee. En als vierde ik denk ook dat LLMs ook bij de detectie van afwijkingen kunnen helpen.

***Interviewer*:**

Ja, klopt, want ik dacht ook dus bij de detectie van afwijkingen daar zijn ze denk ik wel echt super nuttig voor. Maar je hebt natuurlijk ook verschillende soorten LLM's.
Q2.1a: Welke soorten LLMs acht u het meest geschikt voor toepassing in uw sector, en waarom?

Want je hebt bijvoorbeeld GPT-4, Claude, DeepSeek, Apple Intelligence en zoveel meer.

Respondent:
Ik denk dat dat sowieso het fenomeen van de tijd is dat wat je nu een keuze moet maken tussen welk model je wilt gebruiken. Ik denk dat dat in de toekomst weer converteert naar elkaar toe en dat eigenlijk je gewoon een vraag stelt en dat op basis van een eerste assessment door een soort mini LLM de provider zelf besluit naar welk model die het gaat sturen en dat je dat niet meer als gebruiker hoeft te kiezen. Dat is mijn verwachting, maar het hangt er vanaf wat je wil bereiken, hè? Bijvoorbeeld voor het beoordelen van is er momenteel sprake van een afwijking in patronen en detecteren van incidenten. Zou je eigenlijk ***reasoning models*** willen gebruiken en voor dat voorbeeld, wat ik noemde van uitleggen aan gebruikers, dan zou je veel meer een ***general purpose LLM*** voor willen gebruiken, omdat je daar eigenlijk niet hoeft te redeneren, maar meer **v**ersimpelen. Wat ik denk dat het voordeel is van Claude boven ChatGPT is dat Claude een een veel meer context driven , dus dat je daar een grotere context mee kan geven, wat voor complexe cybersecurity problemen zeker een meerwaarde kan zijn.
Maar dat is ook een beetje een soort rat race, want nu is Claude weer degene met de grootste context en dat je er weer op wachten dat ChatGPT weer met een nieuw model of een aanpassing van een bestaand model komt dat daar ook het context window omhoog gaat dus.

***Respondent*:**

Kijk, voor de westerse wereld zou ik nooit naar DeepSeek kijken. Ik denk eigenlijk dat er maar op dit moment twee relevante aanbieders zijn: Anthropic en ChatGPT of eigenlijk openAI. En je ziet dat veel van de andere proposities daarop gebaseerd zijn of zich daar niet mee kunnen meten. GenAi is best grappig, maar kan zich niet meten met de laatste modellen van een Anthropic of van OpenAI.

***Interviewer*:**

En dan een vraag over binnen uw organisatie zelf, want u heeft net gezegd dat u bij 3 organisaties werkt.

FQ2.1a Is er al geëxperimenteerd met de LLMS of ingezet bij bepaalde toepassingen of is dat nog grijs gebied binnen de organisaties?

***Respondent*:**

Nou bij [NAAM BEDRIJF] wordt er nu aangemoedigd om een beetje te experimenteren met de gratis versie van ChatGPT . Dat is grappig, maar ik denk dat dat onvoldoende is, ik denk dat je eigenlijk altijd de betaalde versie moet gebruiken of marktwaarde waar je echt serieuze waarde uit zou kunnen halen. En binnen [NAAM BEDRIJF] zijn er beetje aan het experimenteren met Cursor AI. Dat is met name een een tool op het gebied van software development en AI assisted coding.

Die weer de eigenlijk gewoon de modellen van een anthropic, Google, en OpenAi erachteraan biedt met hun eigen specifieke training om efficiënt AI assistent coding te kunnen doen. Dus daar zijn we een beetje mee aan het experimenteren. Maar het is natuurlijk maar een hele kleine organisatie, dus die leeft met name verder op wat [NAAM BEDRIJF] doet. [NAAM BEDRIJF] Is echt opzicht een aantal stappen verder. Die hebben recent een proof of concept gedaan met ChatGPT Enterprise met 25 medewerkers. En daar is de conclusie dat dat ongeveer 10% efficiency levert gemiddeld per week per medewerker. En dat is natuurlijk gewoon super interessant, want ook al kost die licentie iets, 10% is echt wel significant

En, dat is zonder dat mensen getraind zijn in hoe ze het moeten gebruiken of wat dan ook is. Ook daar is nu al de vraag gesteld of we dat percentage nog verder omhoog kunnen krijgen als er mensen gaan trainen? Nou, dat weet ik niet, dat gaan we zien, maar het is daarmee eigenlijk ook een no brainer om Chat GPT uit te rollen voor de hele organisatie.
Anderzijds hebben we vorig jaar ook al een oproep gedaan met het opnemen van een AI feature in het product wat [NAAM BEDRIJF] levert of een van de producten

die zij leveren en de PoC (proof of concept) was in de basis succesvol. Alleen dan zie je toch wel dat het lastig is om van een werkende PoC naar productie-ready functionaliteit te gaan. Niet alleen op technisch gebied, maar ook zeker op compliance gebied. Omdat er dan toch data naar de externe partijen gaat en aangezien dit ook het matchen van betaalde rekeningen en binnengekomen betalingen ging, heb je daar ook te maken met personal data. En dat is ja gewoon een interessante case.

***Interviewer*:** Ja en [NAAM BEDRIJF] Zijn dan natuurlijk een stapje verder dan [NAAM BEDRIJF].

Q2.2: Hoe wordt in uw organisatie toezicht gehouden op AI-systemen?

Respondent:
Nee, daar is nog relatief weinig voor. Het was wat ik al zei, echt een proof of concept dus Er zijn een paar basis regeltjes van gebruik je verstand en dat soort dingen maar het voordeel natuurlijk van ChatGPT enterprises is dat je niet na hoeft te denken over wat je erin mag stoppen, want het wordt niet gebruikt voor voor training of wat dan ook, want je betaalt gewoon echt ervoor in euro's, dus er hoeven ook veel minder spelregels te zijn over je mag er geen klant data ingooien of je mag dit niet of je mag dat niet.
Waar denk ik nog wel, de uitdaging zit in is hoe de AI ACT dat ook noemt, AI iteracy. Ik denk dat er teveel gebruikers zijn die eigenlijk geen flauw benul hebben hoe je het gebruikt en hoe het werkt.

***Interviewer*:**
Dus dat de educatie daarvan mist, dus dat soort van een kennissessie en educatie cursussen daarvoor moet woord geven.

Respondent:

Ja, precies. Dat is wel handig en eigenlijk noodzakelijk voor elk bedrijf.

***Interviewer*:**
Ja, je had net al compliance opgenomen, maar je hebt natuurlijk ook andere wetgevingen zoals GDPR Dora en de AI ACT.

Q2.3: Hoe beïnvloeden reguleringen zoals GDPR, DORA, NIS2 of de AI Act de toepassing van LLMs in cybersecurity?

***Respondent*:**
Ik denk dat dat afhangt van het gebruik van AI
Wat ik al zei als je kijkt naar het gebruik van AI in het dagelijkse werk, dan heb je eigenlijk relatief weinig te maken DORA en NIS2, GDPR natuurlijk wel maar niet anders dan wat al voor andere systemen die werden gebruikt. Dus dat is ook voor de meeste medewerkers niks nieuws en ze zijn al in getraind. Het gat zit in de AI Act, dat is relatief nieuw. Denk dat veel organisaties daar nog moeten kijken of wat betekent dit nou concreet? Wat moet ik nou echt doen om hieraan te voldoen? Dat soort aspecten, maar ook daar denk ik dat het wordt gebruikt op de werkvloer meevalt, want heel veel mensen zullen ChatGPT toch wel al gebruiken als een zoekmachine.

We hebben het natuurlijk ook vorige week in de AI sessie besproken. Als je nu Google gebruikt, gebruik je op de achtergrond ook AI, dus ik zie dat niet essentieel anders waar compliance een heel grote rol speelt. Dus als je dat in je product gaat aanbieden en jij bent degene die de AI inkoopt hè? Die zeg maar, het contract heeft met de LLM provider, want dan wordt DORA ineens ontzettend relevant, want dan is het een subcontractor. Dan is het een potentiële subverwerker. Al dat soort aspecten van compliance. En dan wordt er eigenlijk ook ineens een stuk relevanter, omdat je dat wat je daar doet ook doorverkoopt aan jouw klanten en dat potentieel daar de klanten van er ook weer plezier van hebben. Als jij het niet helemaal goed doet. Dus ik denk dat het daar veel relevanter is. Ik denk dat er organisaties zijn die goed zijn in compliance en ik denk dat we organisaties zijn die weinig kunnen in compliance.

Interviewer:
Dat is duidelijk, daarvoor had u al iets opgenomen over weinig kennis. Dat is natuurlijk een risico, wat vaak voorkomt bij AI.

Q2.4 Zijn er nog andere ethische risico's ziet u bij de inzet van LLMs? Dus bijvoorbeeld bias en privacy zijn soort van de belangrijkste risico's waar mensen zich zorgen over maken. Maar denkt u zelf dat er ook nog andere risico's zijn?

Respondent:
Kijk wat je natuurlijk altijd wel ziet is hoe meer je in een applicatie automatiseert en ik zie wel wat AI en een potentie heeft om nog weer een extra niveau aan automatisering toe te voegen. Dat dat ook betekent dat er gewoon het personeelsaantal naar beneden moet. Ook in dat voorbeeld, wat ik gaf van die 10% efficiency bij [NAAM BEDRIJF].

Zeggen wel de CEO's van ja hartstikke leuk, maar als we dit gaan uitrollen, willen we ook gewoon een target per afdeling zien van hoeveel personeel gaan eruit of hoeveel vacatures ga je niet vervullen Omdat je nu AI krijgt. Dat is niet zozeer het gevolg puur binnen het LLM zelf, maar wel van AI in zijn algemeenheid. Een ethisch stukje daarvan is in hoeverre moet techniek de banen van mensen vervangen?
En voor sommige banen denk ik dat dat heel goed is en voor anderen misschien wel niet omdat soms de Human touch nodig is en blijft.

Interviewer:

Ja, want dat is dan precies ook mijn volgende vraag, want mijn volgende vraag ging over de menselijke controle.

Q2.4a: Er zijn natuurlijk bepaalde dingen die geautomatiseerd kunnen worden, maar nog steeds wel menselijke controle nodig hebben. En ja, sommige gaven al bij de vorige interviews aan dat er niet alles nog menselijke controle nodig heeft. Hoe denkt u daar zelf over?

Respondent:
Ik denk dat dat uiteindelijk risico gebaseerd moet worden. En ook vooral impact gebaseerd. LLMs kunnen prima aangeven hoe zeker ze zijn van hun antwoord, zeker als je APIs gebruikt. In de API kan je zien in hoeverre die vectoren daadwerkelijk matchen en als dat percentage heel hoog is en de potentiële impact van het verkeerde match relatief klein is. Dan heb ik er geen probleem mee om zonder tussenkomst van mensen dat toe te passen. Dat voorbeeld wat ik gaf daarnet van het matchen van openstaande facturen op binnengekomen betalingen. Het meeste wordt natuurlijk in de klantenkring gewoon per automatische incasso betaald, maar een klein deel schrijft nog gewoon aan het matig de premie van een polis over. Nou, dan moet je daar een polisnummer of een relatienummer in opnemen. Je kunt je voorstellen dat daar toch nog wel mensen zijn die daar een typo in maken of helemaal niks zetten of wat dan ook ja.
Dan moet je kijken naar match, het bedrag, welke naam staat op de rekening, dat soort aspecten nou, dan kun je je afvragen, hoe erg is het als ik bij een premie van 3 tientjes en een waarschijnlijk een match van 99% gewoon zeg oké, ik beschouw die premie als betaald of misschien vervelender misschien niet betaald. Voor je het weet staat er een deurwaarder voor de deur. Ik denk dat je daar altijd vanuit wat de consequentie moet kijken, want er zijn denk ik een andere voorbeelden In de wereld waar die consequenties vele malen groter zijn.

Interviewer:

Nee, top dan had ik ook nog een vraag over de data, want wat ik ook had opgemerkt bij mijn vorige interviews, is dat er een soort van concern was , een soort over synthetische data, dus data die al gegenereerd zijn door LLMs. En dat het voor kan komen dat het als input data kan worden gebruikt.

Q2.5: Hoe belangrijk vindt u de kwaliteit, herkomst en controle van de data die gebruikt wordt in AI-toepassingen zoals LLMs?

Respondent:

Ik denk dat het heel lastig is om dat goed te valideren, want hoe bepaal je wat waar is? Het internet neemt ook razendsnel informatie van elkaar over. Je zou kunnen zeggen: Pas als iets minimaal tien keer voorkomt, beschouw ik het als waar, maar ook dat is geen garantie.

Tegelijkertijd zouden LLM's steeds beter moeten worden in het herkennen van patronen in de data waarmee ze getraind worden. Een model zou moeten kunnen onderscheiden of een tekst afkomstig is van een mens of van een ander LLM. Persoonlijk vind ik het niet per se problematisch als LLM's getraind worden op output van andere modellen, zolang die output van hoge kwaliteit is. Het echte risico zit in het feit dat je een LLM kunt vragen om complete onzin te genereren, en dat die onzin vervolgens in de trainingsdata terechtkomt. Dát is zorgelijk.

Interviewer:

Verder had ik nog een vraag over de controles van het inzetten van een LLM:

Q2.6: Op welke manier wordt in uw organisatie gecontroleerd hoe LLMs worden ingezet, geüpdatet of aangepast? Denk bijvoorbeeld aan prompt filtering etc.

Respondent:

Ik denk dat je inderdaad op de prompt moet controleren. Ik denk ook dat je. eigenlijk op de de respons een bepaalde filtering zouden toepassen als die ongewenst is op het gebied van racisme of andere aspecten.

Dat je die respons niet door laat gaan. Ja, want Ik denk dat je prompt filtering heel sterk kan zijn, maar als je slim genoeg bent, kun je toch een prompt formuleren die door de promptfiltering heen komt die toch per grote onzin of ongewenste reactie schrijft.

Ik denk zeker ook dat je over moet nadenken wie welke modellen gebruiken. Als je alles in applicaties integreert, welke modellen wil ik kunnen gebruiken.

Sommige modellen zijn veel vatbaarder voor het geven van bepaalde responses dan anderen. Dus Je kunt daar ook op het moment dat je het integreert in je applicatie bepaalde keuzes maken om bepaalde modellen gewoon niet in die API ter beschikking te stellen.

Interviewer:
Nou top, de meeste kernvragen zijn al behandeld en sommige had u al beantwoord in een andere vraag. Ik stel voor om nu verder te gaan naar het framework zelf. Heeft u het framework al zelf doorgelezen?

Respondent:

Ja, ik haal hem erbij.

Interviewer:
Nou, een kleine introductie of het framework. Het framework is dus gebaseerd op de literatuuranalyse wat ik hiervoor heb gedaan. Dat bestond dus uit de governance en de compliance wetgevingen. En wat LLMs überhaupt zijn. Daaruit heb ik dus besloten dat de Oranje vakjes staan voor vakjes waar LLMs eigenlijk gelijk toegepast kan worden en de groene is waar het in ieder geval kan helpen, maar het hoeft niet per se gebruikt te worden. Ik heb dus 6 elementen en mijn eerste vraag is dan meteen:

Q2.7: Wat zou u nodig hebben om vertrouwen te hebben in een AI-ondersteund framework voor cybersecurity-risicomanagement?

Respondent:

Om vertrouwen te hebben in zo'n framework is het voor mij belangrijk dat de verschillende componenten helder en begrijpelijk zijn. Sommige termen, zoals 'data quality & information governance', kunnen voor interpretatie vatbaar zijn. Als je daar een duidelijke toelichting bij geeft, draagt dat bij aan de toepasbaarheid en geloofwaardigheid van het geheel. Daarnaast speelt governance een sleutelrol. Een framework kan technisch goed in elkaar zitten, maar als het niet duidelijk is wie waarvoor verantwoordelijk is, of wie toezicht houdt, dan werkt het in de praktijk niet. Voor mij is 'Governance and Oversight' dus het fundament waarop de rest steunt. Zonder dat werkt de implementatie gewoon niet.

Interviewer:
Q2.8: Welke onderdelen vindt u het meest relevant of juist problematisch in uw praktijk? Mist u bepaalde elementen of ziet u mogelijkheden voor verbetering?

***Respondent*:**
Ja, waar valt hier in voor jou het detecteren van afwijkingen in de bestaande maatregelen. Dus eigenlijk het detecteren van incidenten.

Interviewer:
Van incidenten dacht ik zelf dus dat ik dat bij de risk identification kon doen, maar na het interview van hiervoor werd er aanbevolen om een apart element te maken die dan een connectie kon maken met een risk identification.

Respondent:
Ja, dat denk ik ook. Want, want risico's zijn denk ik eigenlijk gewoon iets wat je bedenkt en incidenten zijn eigenlijk het materialiseren van je risico's.

Interviewer:
Dus dat zou naar de Risk Identification gaan. En ook wat u dus hier vooral zei: de educatie. Dus hoe het gebruikt kan worden. Ik kreeg ook een tip om dat te verwerken in het framework. Zelf dacht ik dat dat toegevoegd kan worden aan AI Life Cycle.Heeft u daar verder nog een opmerking over? Het kan in principe ook een apart element zijn.

***Respondent*:**
Kan ook een apart element zijn, het zou ook bij governance kunnen horen.

Interviewer:

En, waarom denkt u dat?

***Respondent*:**

Omdat ik denk dat training een onderdeel is van je cognos model.
Maar ik denk uiteindelijk dat een apart blok logischer is.

Interviewer:
Dat is goed om te weten. Dat waren eigenlijk mijn vragen over het framework. Heeft u zelf nog opmerkingen of aanbevelingen met betrekking tot het gebruik van LLM's in cybersecurity? Of denkt u dat we de belangrijkste punten al hebben behandeld?

Respondent:
Nee ik denk dat we meeste dingen al hebben besproken.

Afsluiting:
Interviewer:
Oké, top. De bedoeling is dat ik de resultaten van dit interview ga verwerken in het framework, dat op basis daarvan dus wat aangepast wordt. In de tweede week van juni stuur ik vervolgens een korte Google-survey naar alle deelnemers van de interviews. Daarbij voeg ik ook de vernieuwde versie van het framework toe, zodat ik die kan valideren.

De survey zal uit ongeveer vijf à zes eenvoudige, gesloten vragen bestaan en ik geef daar direct een deadline bij mee. Ik stuur het begin van de tweede week van juni, met als deadline het einde van diezelfde week.

Daarmee zijn we aan het einde gekomen van dit interview. Nogmaals dank voor uw tijd en waardevolle inzichten, ik heb hier echt veel aan gehad. Zoals eerder aangegeven worden alle antwoorden volledig geanonimiseerd en direct na afronding van mijn scriptie verwijderd. Er wordt niets opgeslagen of gebruikt voor andere doeleinden.

Respondent:

Top, Graag gedaan en veel succes!