Opening Statement

Interviewer:
Goedemorgen. Ten eerste wil ik u bedanken dat u bereid bent deel te nemen aan het interview. Mijn naam is Imane, ik doe momenteel mijn master aan de Universiteit van Amsterdam in de richting Information Studies. Via [Naam] ben ik bij u terechtgekomen. Dit interview maakt deel uit van mijn scriptieonderzoek bij Kouters van der Meer. Het onderzoek gaat over hoe kunstmatige intelligentie effectief kan worden geïntegreerd in het cybersecurity-risicomanagement, met name bij financiële instellingen. Het interview duurt ongeveer 60 minuten, maar denk dat het eerder ongeveer 45 minuten duurt. Alles wordt anoniem verwerkt en er wordt alleen een transcript gemaakt, dat daarna wordt verwijderd.

Het interview bestaat uit twee delen. Eerst is een korte introductie van ongeveer 5 minuten en daarna ga ik naar de kernvragen.

Respondent:
Top dat klinkt goed, ook fijn dat je de briefing had opgestuurd.

Sectie 1: Achtergrond van de geïnterviewde

Interviewer:
Q1.1: Kunt u kort iets over uzelf vertellen en uw rol binnen de organisatie?

Respondent:

Mijn naam is [Naam]. Ik heb gestudeerd in Tilburg en aan de Erasmus Universiteit, met een focus op financiële business, economie en ICT. Daarna ben ik verhuisd naar Utrecht waar ik nu woon met mijn vriendin en twee kinderen. En in Utrecht heb ik mijn RE-opleiding afgerond. Vervolgens heb ik ongeveer zes jaar gewerkt bij [Naam bedrijf]. Na die periode heb ik de stap gezet naar het ondernemerschap. Inmiddels leid ik een team van tien mensen dat zich richt op red teaming, pentesting, social engineering en gerelateerde security testen. We kruipen als het ware in de huid van de aanvaller om organisaties beter te beschermen. Ons ontwikkelteam maakt daarnaast ook -oplossingen, waarbij we AI gebruiken als input bron voor intelligence en bevindingen en het model zelf draait niet op AI, maar we gebruiken het ter ondersteuning van onze analyses.

Interviewer:
Q1.2: Hoeveel ervaring heeft u met cybersecurity, risicomanagement of AI?

Respondent:

Ja, ik heb inmiddels behoorlijk wat ervaring met cybersecurity. Ik zit al zo'n 18 jaar in het vak, begonnen in de tijd dat ik bij [Naam Bedrijf] en later bij [Naam Bedrijf] werkte. Sindsdien is het altijd een onderdeel van mijn werk geweest. Wat betreft AI is mijn ervaring wat recenter, logisch ook gezien de ontwikkeling van de technologie. Maar inmiddels doen we er veel mee , onder andere met tools als ChatGPT en Claude, die we onderling vergelijken. We hebben hier zelfs een eigen LLM draaien als testomgeving. Daarnaast zijn we betrokken bij een WBSO-project, gesubsidieerd via de RVO, waarmee we AI verder verkennen en ontwikkelen binnen onze organisatie.

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

Mijn ervaring met AI is relatief recent, wat ook logisch is gezien de nog jonge aard van de technologie. In onze organisatie werken we inmiddels wel actief met tools zoals Claude en ChatGPT . Daarnaast hebben we een eigen Large Language Model opgezet als testomgeving, waarmee we experimenteren. Dit project wordt deels ondersteund via een WBSO-subsidie van de RVO. Hoewel ik mezelf geen AI-expert zou noemen, ben ik zeker bekend met LLMs en zie ik hoe ze toegepast kunnen worden binnen ons werk, bijvoorbeeld bij het genereren van inzichten en ondersteunen van red teaming activiteiten.

Interviewer:
Oke top dus u weet wel wat de Large Language Models zijn. Dat is fijn, want dat is eigenlijk de AI-technologie die ik dan gebruik.

Dat waren de introductie vragen. De volgende vragen zijn meer inhoudelijk en over het framework.

Sectie 2: Thematische verdieping

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

Respondent:

Ik zie dat LLMs met name waarde kunnen toevoegen op drie gebieden. Ten eerste zijn ze goed in het leggen van relaties tussen concepten en tokens, ze kunnen verbanden herkennen binnen complexe informatieclusters. Dat maakt ze krachtig bij het analyseren van risico's en het koppelen van die risico's aan bijvoorbeeld andere

controleframeworks of bestaande databronnen. Ten tweede kunnen ze helpen bij het genereren van voorspellingen, bijvoorbeeld door op basis van eerder waargenomen patronen aanvullende risico's aan te dragen. En als laatst zijn LLMs sterk in het creëren van teksten. Dat maakt ze geschikt voor het automatisch opstellen van rapportages, toelichtingen of aanbevelingen binnen risicomanagement processen. De echte meerwaarde zit voor mij vooral in de inhoudelijke ondersteuning, patroonherkenning en tekstgeneratie.

Interviewer:
U benoemde in de introductie al twee soorten tools, namelijk Claude en ChatGPT .
Q2.1a: Zijn er nog andere soorten LLMs die u het meest geschikt vind voor toepassing in uw sector, en waarom?

Respondent:

Wij gebruiken momenteel Claude, en persoonlijk vind ik die beter dan ChatGPT . Dat komt vooral doordat Claude vaak relevantere en inhoudelijk sterkere antwoorden geeft. Natuurlijk hangt dat ook deels af van hoe je de prompt formuleert, maar in onze ervaring is Claude consistenter. Daarnaast zijn we bezig met het ontwikkelen van een eigen LLM. Dat model is gebaseerd op een open-source LLM, waar we vervolgens eigen functionaliteiten aan hebben toegevoegd. Toch moet ik eerlijk zeggen dat ons eigen model (nog) niet in de buurt komt van wat Claude of ChatGPT kan. De prestaties van de grote, commerciële modellen blijven vooralsnog superieur.

Interviewer:
Duidelijk, interessant dat jullie zowel met Claude werken als een eigen model ontwikkelen. U gaf ook aan dat jullie zelf functionaliteiten hebben toegevoegd op basis van een open-source model. Dat roept bij mij de vraag op hoe jullie de inzet van die AI-toepassingen organiseren en beheersen.

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

Respondent:

Binnen onze organisatie hanteren we de duidelijke afspraak dat gevoelige data nooit mag worden ingevoerd in publieke AI-systemen. Om dat te waarborgen, werken we standaard met placeholders. Dus in plaats van bijvoorbeeld klantnamen gebruiken we termen zoals 'company' of 'friendname', en andere gevoelige informatie wordt op dezelfde manier geanonimiseerd. Als het echt noodzakelijk is om met gevoelige data te werken, dan gebruiken we onze eigen LLM. Die is op een veilige manier geïmplementeerd en werkt onder andere via het principe van MCPs , misschien ken je dat? Dat zijn als het ware koppelvlakken waarmee je een LLM kunt laten

communiceren met andere LLMs of diensten. Daar zijn we nu mee aan het experimenteren, zodat we gecontroleerd met gevoelige of waardevolle data kunnen werken binnen een afgeschermde omgeving. In sommige gevallen gebruiken we dus onze eigen LLM in combinatie met een extern systeem, maar altijd via een veilige route. Ons eigen model draait bovendien volledig geïsoleerd op netwerkniveau. Het heeft namelijk geen internettoegang en is van buitenaf niet benaderbaar. We proberen het gebruik van gevoelige data sowieso zoveel mogelijk te beperken, tenzij het echt nodig is om effect te behalen. Maar zelfs dan nemen we technische en organisatorische maatregelen om dat veilig te doen.

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

Respondent:

Als je het vanuit geopolitiek perspectief bekijkt, zie je duidelijke verschillen. In Azië en de Verenigde Staten wordt er aanzienlijk minder belang gehecht aan privacy dan hier in Europa. Daardoor kunnen ontwikkelingen op het gebied van AI veel sneller plaatsvinden. Er is meer ruimte voor experiment en grootschalige toepassing, terwijl wij hier gebonden zijn aan strengere wet- en regelgeving zoals de GDPR en de AI Act. Dat leidt ertoe dat innovatie in Europa langzamer verloopt. Vergelijk het met de ontdekking van koolstof: als het er eenmaal is, kun je het niet meer 'ontdekken'. Zo is het ook met AI, je kunt niet ineens besluiten het niet meer te trainen of in te zetten voor bepaalde toepassingen. Die ontwikkeling gaat sowieso door, ongeacht regelgeving.

Dat gezegd hebbende, begrijp ik de reden achter die regelgeving heel goed. Het is terecht dat we in Europa proberen AI verantwoord en transparant te reguleren. Maar het heeft wel als consequentie dat onze innovatiekracht mogelijk achterblijft ten opzichte van regio’s waar de regels minder streng zijn.

Interviewer:
Helder, u maakt een goed punt over hoe regelgeving innovatie kan afremmen, maar ook noodzakelijk is om AI in goede banen te leiden. Dat brengt me eigenlijk meteen bij een andere belangrijke dimensie van verantwoord AI-gebruik, namelijk de ethische kant ervan.

Q2.4: Welke ethische risico's ziet u bij de inzet van LLMs, bijvoorbeeld rond bias, discriminatie of foutieve classificatie?

Respondent:

Ja, daar zit zeker iets in. Kijk, een LLM doet uiteindelijk gewoon wat je het leert. Je stopt er menselijke logica of gedrag in zodat het lijkt alsof het kan redeneren, maar het blijft een weerspiegeling van de data waarop het is getraind. Als die data vooral komt van bijvoorbeeld witte westerse auteurs, dan zit dat perspectief automatisch ingebakken in het model. Wat je nu ook steeds vaker ziet, is dat AI nieuwe content genereert op basis van eerdere AI-content. Je krijgt dan een soort self-fulfilling loop: oude data voedt nieuwe output, en die nieuwe output wordt weer als input gebruikt. Zo versterk je bepaalde aannames of blinde vlekken zonder dat je het misschien door hebt. En dat gaat niet alleen over etniciteit of huidskleur, maar ook over culturele achtergrond, seksuele geaardheid of geografische bias. De westerse wereld is simpelweg oververtegenwoordigd op internet, dus landen als Botswana komen dan nauwelijks terug in de datasets, terwijl dat ook waardevolle perspectieven zijn. En eerlijk gezegd maak ik me ook zorgen over hoe mensen ermee omgaan. Soms zie ik developers die klakkeloos code uit ChatGPT kopiëren en in hun systeem plakken zonder echt te begrijpen wat het doet. Je moet blijven nadenken. AI is een handig hulpmiddel, maar het moet je denkproces niet vervangen. Schrijf af en toe gewoon zelf die management summary, en denk kritisch na over wat een model je aanreikt. Gebruik het slim, maar blijf zélf sturen.

Interviewer:
Okay top u noemde al op dat mensen moeten blijven na denken. En dat is ook meteen waar mijn volgende vraag over gaat, namelijk:

Q4.a: Vindt u dat er altijd menselijke controle nodig is, of kunnen bepaalde beslissingen volledig geautomatiseerd worden?

Respondent:
Nou ja, ik denk dat als je AI echt effectief wilt inzetten, je zo min mogelijk controlemomenten moet inbouwen. Je zou alleen controle moeten toepassen waar het echt nodig is. Dat hangt sterk af van je bedrijfsvoering, je beleid en van wat je wel of niet als 'actionable' beschouwt. In het begin zullen er waarschijnlijk veel controles zijn, gewoon omdat het nieuw is en mensen nog vertrouwen moeten opbouwen. Maar naarmate de tijd vordert en het vertrouwen groeit, zullen die controles vanzelf minder worden. Dus ja, zeker controleren waar nodig, maar niet overcompenseren.

Interviewer:
Dus wat ik ervan begrijp is dat menselijke controle sowieso nodig is maar het moet niet te veel worden gebruikt. Oke, dat is helder. Verder had u ook al wat opgenomen over de data input:

Q2.5: Hoe belangrijk vindt u de kwaliteit, herkomst en controle van de data die gebruikt wordt in AI-toepassingen zoals LLMs? En welke aanpak hanteert uw organisatie daarin?

Respondent:
De kwaliteit en herkomst van data zijn absoluut belangrijk, juist omdat we er in de praktijk vaak weinig directe controle over hebben. Binnen onze organisatie proberen we dat te ondervangen door modellen in een gecontroleerde, afgeschermde omgeving te draaien en waar mogelijk input data te structureren of te filteren. Maar echt diepgaand controleren op bronniveau is meestal niet haalbaar. Dat heeft vooral te maken met schaal en middelen. Zeker kleinere organisaties of start-ups hebben vaak niet de capaciteit om een volledig eigen LLM te trainen vanaf nul. Het kost simpelweg te veel tijd, geld en rekenkracht. Daarom bouwen de meeste partijen voort op bestaande open-source modellen, zoals DeepSeek of andere foundation models, en ontwikkelen ze daar hun eigen toepassingen bovenop.

Het nadeel daarvan is dat we er met z'n allen een soort blind vertrouwen in hebben ontwikkeld. We nemen aan dat die modellen goed zijn getraind, en als er dan iets geks uitkomt, realiseren we ons vaak pas later dat we zelf nauwelijks zicht hebben op de oorspronkelijke trainingsdata. Je kunt wel tokens en relaties zien, maar niet de precieze bron inhoud. We vertrouwen op de reputatie van de maker, maar echte transparantie ontbreekt. Ethisch en wetenschappelijk gezien zou je eigenlijk je eigen modellen moeten trainen. Maar dat is voor de meeste organisaties simpelweg niet realistisch. Kijk bijvoorbeeld naar hoe lang het heeft geduurd om GPT te ontwikkelen, dat is onhaalbaar voor de meeste bedrijven.

Wat betreft specifieke modellen zoals DeepSeek zie je dat de meningen verdeeld zijn. Sommige mensen zijn terughoudend vanwege de herkomst (China), anderen gebruiken het zonder bezwaar. Mijn visie is: zolang je het model isoleert in een eigen, veilige omgeving zonder internettoegang, kun je het verantwoord inzetten, mits je bewust bent van de bredere geopolitieke context. Die context wordt sowieso steeds belangrijker. Tijdens de Cybersec-beurs in Brussel merkte ik al dat geopolitieke voorkeuren ook keuzes voor AI-modellen, software en cloudleveranciers steeds meer beïnvloeden. Dat gaat de komende jaren alleen maar toenemen.

Interviewer:

Ja, helder. Het lijkt ook wel logisch dat veel organisaties vertrouwen op bestaande modellen aangezien bijna iedereen het al gebruikt. Maar dan loop je natuurlijk tegen de transparantie over de herkomst van de data. Dat brengt me meteen bij een volgende onderwerp: de manier waarop die modellen worden beheerd en gecontroleerd binnen de organisatie.

Q2.6: Op welke manier wordt in uw organisatie gecontroleerd hoe LLMs worden ingezet, geüpdatet of aangepast? En zijn er afspraken over welke modellen wanneer gebruikt mogen worden?

Respondent:

Ja, kijk, we houden dat gewoon strak in de gaten. Alles wat met LLMs gebeurt, doen we in een afgesloten omgeving. Dus als er iets raars gebeurt, dan blijft het daarbinnen. Het draait gewoon los van de rest, zodat het geen gekke dingen kan doen in het primaire proces. We testen eigenlijk altijd eerst: wat doet het model, waar zitten de risico's, hoe reageert het op bepaalde prompts. Dus wij geven altijd gewoon input, zeg maar. Wij gebruiken het bijvoorbeeld voor het schrijven van vindingen. Dat scheelt gewoon ontzettend veel tijd. Dus als hij in één keer daar hele rare dingen in gaat zetten, dan zien wij dat nog voordat dat naar een grotere klant gaat.. En als het niet goed voelt, dan gaat het niet live. Maar er gaat sowieso altijd een menselijke check over. Nooit automatisch naar buiten of richting klant.

Qua modellen hebben we wel wat afspraken. Als het gevoelige data is, dan gebruiken we óf placeholders óf onze eigen LLM. En die is ook weer volledig afgeschermd. We hebben dat dan gekoppeld via MCP's. Dus ja, we hebben geen formeel beleidsdocument met tien lagen governance, maar in de praktijk zit er gewoon een goed werkend systeem achter. We doen niks op de gok, alles is afgeschermd, getest, en er zit altijd een mens bovenop.

Interviewer:
Dat waren eigenlijk alle kernvragen. U had een paar vragen al in beantwoord in een andere vraag dus die ga ik uiteraard niet opnieuw stellen. De volgende vragen gaan over het framework:

Q2.7: Wat zou u nodig hebben om vertrouwen te hebben in een AI-ondersteund framework voor cybersecurity-risicomanagement? En hoe zou zo'n framework idealiter getoetst of verbeterd moeten worden in de praktijk?

Voor mij moet een framework praktisch toepasbaar zijn en aansluiten op hoe organisaties nu al met AI werken. Dus niet alleen op hoog abstractieniveau blijven hangen, maar ook echt laten zien welke technische maatregelen je kunt nemen.

Denk aan isolatie van modellen, het gebruik van audit trails, en bijvoorbeeld maatregelen zoals MCP's waarmee je veilig kunt koppelen tussen systemen. Daarnaast moet het framework ook mee kunnen bewegen met technologische en geopolitieke ontwikkelingen. Je ziet dat er steeds sneller veranderingen komen in hoe we AI gebruiken en hoe regelgeving zich ontwikkelt. Een framework moet dus flexibel genoeg zijn om dat bij te houden. Wat ook helpt voor vertrouwen is als het framework getest is in meerdere contexten: zowel bij grotere banken als bij kleinere organisaties. En dat er gebruik wordt gemaakt van realistische scenario's, zodat je ziet wat er echt gebeurt bij bijvoorbeeld prompt injectie, misclassificatie of datalekken via input.

Interviewer:
Oke top dus wat ik hieruit haal is dat een framework sowieso moet worden getest bij kleine en grote bedrijven. En dat het mee moet kunnen met de tijd, dus alle veranderingen die er plaats zullen vinden. De volgende vraag gaat eigenlijk over het voorlopige framework zelf.

Q2.8: Na het doornemen van het voorlopige framework: welke onderdelen vindt u het meest relevant of juist problematisch in uw praktijk? Mist u bepaalde elementen of ziet u mogelijkheden voor verbetering?

Respondent:

Even kijken hoor. Ja, het hangt er natuurlijk vanaf hoe je het inzet. In jouw framework zie ik het technische aspect wel een beetje terugkomen, rechtsonder bijvoorbeeld, maar dat blijft nog vrij algemeen. Vanuit mijn ervaring, en ik kijk er toch ook met een technische bril naar, merk ik dat de technische kant van AI vaak onderbelicht blijft in dit soort discussies. Terwijl daar juist veel praktische maatregelen mogelijk zijn. Denk bijvoorbeeld aan het isoleren van modellen, of aan het gebruik van MCP’s. Zulke maatregelen zijn enorm waardevol: ze beschermen de vertrouwelijkheid van data én maken AI inzetbaarder in gevoelige omgevingen.

Ik zeg niet dat je daar meteen een heel nieuw framework onderdeel van moet maken, maar het zou goed zijn om dit expliciet te benoemen. Een degelijke technische implementatie ondersteunt namelijk direct andere pijlers van je framework, zoals compliance, ethiek en governance. Als je het technisch goed neerzet, voorkom je dat je later allerlei risico’s moet wegredeneren. Je zou dit kunnen verwerken onder ‘AI lifecycle and LLM-specific controls’, zodat het wel meegenomen wordt, zonder het onnodig complex te maken. Wat ik daarnaast mis, is het expliciet benoemen van risk detection als afzonderlijk component. In de praktijk is dat vaak één van de eerste toepassingen van AI in cybersecurity: het signaleren van afwijkingen, anomalieën, ongebruikelijk gedrag. Dat komt nu wat impliciet terug, maar verdient wat mij betreft echt een eigen plek in het framework. Niet alleen omdat het relevant is, maar ook omdat het unieke uitdagingen met zich meebrengt, denk aan false positives, uitlegbaarheid en impact op incident respons AI met al vind ik het

framework inhoudelijk sterk, maar met deze aanvullingen wordt het nog praktischer en beter toepasbaar in de dagelijkse realiteit.

Afsluiting

Interviewer:
Dat brengt ons ook meteen aan het einde van het interview. Nogmaals bedankt voor uw tijd en het delen van uw inzichten, het is ontzettend waardevol voor mijn onderzoek. Zoals eerder aangegeven, worden alle antwoorden volledig geanonimiseerd en vertrouwelijk behandeld. De opname gebruik ik alleen voor het uitwerken van het transcript en de analyse in mijn scriptie. Daarna wordt alles verwijderd. In de tweede week van juni kunt u nog een mail van mij ontvangen met een Google Survey. Dit is nodig voor de validatie van het final framework en zal dan gelijk ook de deadline meesturen.

Heeft u verder nog vragen of opmerkingen?

Respondent:
Ja, wanneer moet je het inleveren?

Interviewer:
Goede vraag, mijn deadline is 27 juni en een week daarvoor stuur ik mijn final draft op. Want dan heb ik nog genoeg tijd om iets aan te passen