

VERKOELING VAN WERKDRUK MET AI IN DE KOELSECTOR

Een onderzoek naar de inzet van
Large Language Models voor de
ondersteuning van
koelsysteemmonteurs

VERTROUWELIJK



Maxim van Duin (500853661)
Hogeschool van Amsterdam
Toegepaste Wiskunde
12 juni 2024

Vertrouwelijk

Verkoeling van werkdruk met AI in de koelsector

Een onderzoek naar de inzet van Large Language Models voor de ondersteuning van koelsysteemmonteurs

Auteur: Maxim van Duin

Studentnummer: 500853661

Bedrijf: FreezerData

Afdeling: R&D

Bedrijfsbegeleider 1: Daniël Overdevest

Functie: CTO

Bedrijfsbegeleider 2: Henk Komin

Functie: Directeur Onderzoek en Ontwikkeling

Onderwijsinstelling: Hogeschool van Amsterdam

Faculteit: Techniek

Opleiding: Toegepaste Wiskunde

Docentbegeleiders: Robin Buijs en Michelangelo Vargas Rivera

Tweede examinator: Robert Nolet

Datum: 12-06-2024

Managementsamenvatting

Dit rapport presenteert een onderzoek naar de inzet van een *Large Language Model* (LLM) binnen de koeltechnische sector, met focus op de bedrijven IJskoud en FreezerData. Het onderzoek richt zich op hoe een LLM kan worden ingezet om de vragen van monteurs te beantwoorden op basis van koelsysteemdata en handleidingen.

Het onderzoek heeft als doel de werkdruk van de monteurs bij IJskoud te verlichten door gebruik te maken van een LLM. Hiervoor is een automatische data-analyse ontwikkeld die de benodigde informatie uit de data haalt, zoals afwijkingen in de temperatuur van bepaalde leidingen en de run time van de compressor. De run time is de tijd dat de compressor per keer aanstaat. Daarnaast is de basis gelegd voor een *Retrieval-Augmented Generation* (RAG)-framework dat relevante informatie uit de handleidingen ophaalt. Het resultaat is een LLM dat vragen van monteurs kan beantwoorden op basis van data en handleidingen.

De evaluatie van de output van het LLM toont aan dat de gegenereerde antwoorden over het algemeen voldoende en relevant zijn, met weinig tot geen hallucinaties. Hoewel de antwoorden in sommige gevallen als te algemeen worden beschouwd, vooral wanneer de algoritmes van FreezerData geen output hadden, is laten zien dat de inzet van een LLM een veelbelovende oplossing kan zijn voor de vermindering van de werkdruk bij monteurs bij IJskoud.

Ondanks de veelbelovende resultaten, zijn er enkele beperkingen:

- In het huidige proces wordt de informatie uit de data van het koelsysteem nooit gecombineerd met de informatie uit de handleidingen.
- De output van het LLM is op een beperkt aantal storing- en niet-storingssituaties geëvalueerd. In deze situaties beoordeelde het LLM altijd correct of er een probleem aanwezig was. Er kan echter niet gegarandeerd worden dat het LLM altijd juist beoordeelt of er een storing aanwezig is.
- Het LLM is op een beperkt aantal vragen geëvalueerd. De antwoorden hadden soms beperkingen, maar werden over het algemeen als goed beschouwd. Er kan echter niet gegarandeerd worden dat het LLM andere vragen ook juist beantwoordt.

Op basis van de bevindingen in dit onderzoek worden de volgende aanbevelingen gedaan aan FreezerData:

- Er wordt aanbevolen om het LLM in eerste instantie in productie te brengen, waarbij gebruikers alleen de mogelijkheid hebben om een aantal vooraf bepaalde vragen te stellen. Er kan immers niet gegarandeerd worden dat het LLM correcte antwoorden zal geven op andere vragen.
- Er wordt aanbevolen om de output van het LLM uitgebreider te evalueren en hier indien nodig monteurs bij te betrekken.
- Momenteel worden vragen door het model uitsluitend beantwoord op basis van de data-analyse of de handleidingen. Er wordt aanbevolen het proces te verbeteren, zodat het LLM de informatie uit de data-analyse en de handleidingen kan combineren.
- Er wordt aanbevolen om de automatische data-analyse in de toekomst uit te breiden met nieuwe algoritmes en AI-modellen van FreezerData.
- Er wordt aangeraden om het verder ontwikkelde RAG-framework van de UvA-student te gebruiken.
- Er wordt aanbevolen om geheugen toe te voegen aan het LLM voor vloeiendere gesprekken, vooral als er een chatbot van wordt gemaakt.
- Overweeg in de toekomst een ander LLM te gebruiken, maar houd rekening met mogelijke verschillen in reacties op dezelfde prompt.

Inhoudsopgave

1 Inleiding	1
2 Theoretisch kader	3
2.1 Large Language Model	3
2.2 Retrieval Augmented Generation (RAG).....	4
2.2.1 Vectordatabase en embedding.....	4
2.2.2 Prompt.....	4
2.3 Beoordelen LLM en RAG.....	5
2.3.1 Faithfulness.....	5
2.3.2 Answer relevance	6
2.3.3 Answer correctness	6
2.4 Chains en agents.....	7
2.5 Koelsysteem.....	9
3 Data	10
3.1 Dataverkenning.....	10
3.1.1 Koelproces tijdens normale omstandigheden	11
3.1.2 Koelproces tijdens een storingen	12
3.2 Fingerprint	13
4 Methode	14
4.1 Automatische data-analyse	14
4.1.1 Vergelijking van de gemiddelde maximale temperaturen.....	15
4.1.2 Vergelijking van de gemiddelde run time	16
4.1.3 Vergelijking van het gemiddeld aantal schakelingen.....	16
4.1.4 Detectie van trend in de temperatuur.....	17
4.1.5 Detectie van verandering in de temperatuur	17
4.1.6 Overige onderdelen automatische data-analyse.....	17
4.2 Implementatie van het RAG-framework	18
4.3 Proces en prompt	18
4.3.1 Opbouw proces	18
4.3.2 Opbouw prompt	19
4.3.3 Problemen met proces	22
4.3.4 Verbetering proces en prompt	23
4.3.5 Afwegingen nieuw proces	27
4.4 Large Language model.....	27
5 Evalueren en analyseren output LLM	28

6 Conclusie	32
7 Discussie	34
8 Aanbevelingen	36
Literatuurlijst	37
Bijlage A: volledige prompts	40
Bijlage B: ingezoomde schematische weergaven van acties	48
Bijlage C: output van LLM	52
Bijlage D: schematische weergave van uitgebreid proces.....	65

1 Inleiding

De technische sector in Nederland kampt met een groot personeelstekort. In 2023 werkten er 934.000 mensen in de industrie, wat neerkomt op 8% van alle banen in Nederland ('Tekort aan ICT'ers, technici en monteurs die industrie moeten verduurzamen', 2023). Desondanks stonden er in het tweede kwartaal van 2022 maar liefst 86.250 vacatures open in de techniek ('Aanpak personeelstekort in techniek en ICT', 2023).

Daarnaast zorgt vergrijzing voor extra personeelstekort in deze sector: 13% van de werknemers in de technische sector is 60 jaar of ouder, tegenover 10% in andere sectoren. Bovendien is 28% van de werknemers tussen de 50 en 60 jaar oud ('Tekort aan ICT'ers, technici en monteurs die industrie moeten verduurzamen', 2023). Het is waarschijnlijk dat het personeelstekort in de technische sector de komende jaren verder zal stijgen.

Bij IJskoud, een bedrijf gespecialiseerd in koelinstallaties, zagen ze dit probleem aankomen en hebben ze daarom enkele jaren geleden FreezerData opgericht. FreezerData is ontwikkeld met de expertise en ervaring van IJskoud en heeft geresulteerd in algoritmes voor het nauwkeurig voorspellen van storingen bij koelinstallaties. Hierdoor kunnen monteurs zich richten op belangrijk werk: het onderhouden van installaties die daadwerkelijk onderhoud nodig hebben en het installeren van nieuwe systemen.

Ondanks de inzet van FreezerData blijft het personeelstekort een probleem bij IJskoud. Als gevolg van het personeelstekort ervaren monteurs tijdsdruk bij het uitvoeren van hun werkzaamheden. Bij IJskoud en FreezerData willen ze dit probleem oplossen door de efficiëntie en effectiviteit van de monteurs verder te verbeteren. Daarom overwegen ze de inzet van een *Large Language Model* (LLM) dat monteurs kan ondersteunen bij het uitvoeren van hun werkzaamheden.

Een LLM is een type generatieve AI dat staat om teksten te genereren op basis van de input van de gebruiker. Door monteurs toegang te geven tot een LLM, kunnen zij snel en accuraat antwoorden krijgen op hun vragen over het koelsysteem. Dit kan leiden tot een verkorting van de benodigde tijd voor werkzaamheden, omdat monteurs minder tijd kwijt zijn aan het zoeken naar informatie en het interpreteren van handleidingen. Daarnaast kan een LLM helpen bij het interpreteren van de koelsysteemdata, waardoor sneller de oorzaak van een mogelijk probleem bij het koelsysteem kan worden geïdentificeerd. Wat bijdraagt aan een lagere werkdruk en een efficiëntere inzet van het beschikbare personeel.

In dit rapport wordt onderzocht hoe een LLM, op basis van koelsysteemdata en handleidingen, vragen van monteurs over het koelsysteem kan beantwoorden. Deze hoofdopdracht wordt onderverdeeld in de volgende deelopdrachten:

- **Het ontwikkelen van een automatische data-analyse** om belangrijke informatie uit de data te halen.
- **Het ontwikkelen van een basis *Retrieval-Augmented generation (RAG)*-framework** om relevante informatie uit de handleidingen op te halen.
- **Het opstellen van een prompt**, die de basis is voor de input van het LLM.
- **Het ontwikkelen van een proces dat bepaalt welke informatie in de prompt wordt verwerkt**, zodat het LLM over de informatie beschikt die het nodig heeft om de vraag te beantwoorden.
- **Het evalueren van de output van het LLM** om de kwaliteit van de antwoorden te bepalen.

Het doel van het implementeren van een LLM is de benodigde tijd van werkzaamheden te verkorten, wat mogelijk zal leiden tot een vermindering van het personeelstekort en werkdruk bij de monteurs.

Om een LLM te ontwikkelen dat op basis van koelsysteemdata en handleidingen vragen kan beantwoorden, wordt eerst een automatische data-analyse ontwikkeld die belangrijke informatie uit de data haalt voor in de prompt. De prompt is de input van het LLM. Daarnaast wordt de basis gelegd voor *Retrieval-Augmented Generation (RAG)*, waarmee relevante informatie uit de handleiding wordt opgehaald. Vervolgens wordt er een prompt ontwikkeld die onder andere de instructies, de resultaten van de automatische data-analyse en/of de relevante informatie uit de handleidingen bevat. Ten slotte wordt de output van het LLM beoordeeld om de kwaliteit te bepalen.

In hoofdstuk 2 wordt het theoretisch kader opgesteld, waarbij er onder andere literatuuronderzoek is gedaan naar LLMs, RAG en het beoordelen van de output van een LLM. In hoofdstuk 3 wordt de data die in dit onderzoek wordt gebruikt verkend. In hoofdstuk 4 staat de ontwikkeling van de automatische data-analyse en de prompt centraal. In hoofdstuk 5 wordt de output van het LLM geëvalueerd. Hoofdstuk 6 gaat in op de conclusie. In hoofdstuk 7 wordt de discussie besproken. Tot slot wordt er in hoofdstuk 8 advies gegeven aan FreezerData.

2 Theoretisch kader

Om het praktijkonderzoek met voldoende kennis uit te kunnen voeren, is er vooraf een literatuuronderzoek gedaan naar verschillende onderwerpen. In paragraaf 2.1 wordt de werking van een *Large Language model* (LLM) besproken. Om een LLM up-to-date te houden met actuele of domeinspecifieke kennis, kan *Retrieval Augmented Generation* (RAG) gebruikt worden, dit wordt verder uitgelegd in paragraaf 2.2. Hierbij wordt ingegaan op de vectordatabase, *embedding* en de prompt. Vervolgens wordt in paragraaf 2.3 uitgelegd hoe de output van een LLM beoordeeld kan worden, waarbij wordt ingegaan op drie verschillende prestatiemaatstaven. In paragraaf 2.4 worden *chains* en *agents* besproken. In paragraaf 2.5 wordt uitleg gegeven over de werking van een koelsysteem.

2.1 Large Language Model

Large language models zijn een type generatieve AI die in staat zijn om teksten te genereren. Generatieve AI kan naast tekst ook afbeeldingen, video's en andere gegevens genereren (Brynjolfsson, Li, & Raymond, 2023). LLMs voorspellen het volgende woord op basis van de voorgaande woorden (Stöffelbauer, 2023). Deze modellen worden getraind op grote hoeveelheden tekstdata, waarvan het internet de grootste bron is. LLMs zijn meestal gebaseerd op transformers, een type neurale netwerk ontwikkeld door Google. Deze modellen zijn zeer groot, waarbij het aantal parameters kan oplopen tot ver in de miljarden (Luna, 2023). Dankzij de grote hoeveelheid traindata en hoge aantal parameters kunnen LLMs veel taalgerelateerde taken uitvoeren, waaronder het genereren van tekst, vragen beantwoorden, vertalen en samenvatten.

Er zitten ook enkele nadelen aan het gebruik van LLMs. De grootte van het model zorgt ervoor dat het onmogelijk wordt de redenering van het model te achterhalen (Luna, 2023). Daarnaast kunnen LLMs bevoordeeld zijn door onder andere de data waarop het model getraind is. Als deze data niet representatief is voor de hele populatie, zal het model mogelijk niet voor iedereen een passend antwoord genereren (Gallegos et al., 2024). Ook is er het risico op hallucinaties, waarbij het LLM onwaarheden presenteert als de waarheid (Kerner, z.d.).

2.2 Retrieval Augmented Generation (RAG)

Een probleem bij een LLM is het up-to-date houden van actuele kennis of domeinspecifieke kennis. Dit probleem kan op verschillende manieren opgelost worden, bijvoorbeeld door het constant blijven trainen van het LLM op grote tekstdatasets of door RAG te gebruiken. RAG is een techniek waarbij het LLM toegang krijgt tot externe informatiebronnen, hierdoor heeft het LLM toegang tot relevante, betrouwbare en actuele informatie (Barnett, Kurniawan, Thudumu, Brannelly, & Abdelrazek, 2024).

2.2.1 Vectordatabase en embedding

Er zijn verschillende manieren om RAG toe te passen. Om RAG te kunnen gebruiken is er een vectordatabase nodig. Een vectordatabase is een database die hogedimensionale vectoren opslaat. De vectoren zijn representaties van in dit geval tekst. Tekst die vergelijkbaar is, wordt opgeslagen in vectoren die een vergelijkbare representatie hebben (Han, Liu, & Wang, 2023). De informatie die in de vectordatabase moet worden opgeslagen wordt meestal geleverd als pdf-document. Deze pdf-documenten worden omgezet in kleinere stukken tekst. Deze stukken tekst worden *chunks* genoemd. De *chunks* worden met een *embedding*-model omgezet in een vectorrepresentatie (Gao et al., 2024). Deze techniek kan rekening houden met de betekenis en de grammaticale vorm van de zinnen (Wang, Wang, Chen, Wang, & Kuo, 2019). In het proces wordt er ook een index gecreëerd die tekst *chunks* en de vectorrepresentaties opslaat als een paar zodat het makkelijk opgezocht kan worden (Gao et al., 2024). Vervolgens wordt op de *query*, oftewel de vraag van de gebruiker, hetzelfde *embedding* model toegepast. De vector van de *query* wordt vergeleken met de vectorrepresentatie uit de vectordatabase. Het systeem haalt de tekst *chunks* op van de vectoren die het meest overeenkomen met de *query* vector (Gao et al., 2024).

2.2.2 Prompt

De originele *query* van de gebruiker wordt samen met de geselecteerde *chunks* uit de vectordatabase toegevoegd aan de prompt. Een prompt is de input van het LLM. Het LLM heeft de taak om een antwoord te formuleren op de prompt (Gao et al., 2024). Over het algemeen zijn prompts dynamisch en bevatten onder andere de invoer van de gebruiker. Vaak worden er enkele voorbeeldvragen met daarbij het antwoord meegegeven aan het taalmodel om een beter antwoord te genereren (Topsakal & Akinci, 2023). De prompt is een belangrijk onderdeel bij een LLM. Een slechte prompt betekent meestal ook een slechte output. Je kunt zelf bepalen hoe een prompt voor een LLM er uit ziet, maar over het algemeen bevat een prompt de volgende onderdelen: instructies, externe informatie, input gebruiker en output indicator (Briggs & Ingham, z.d.).

- Instructies: vertelt het model wat het moet doen, hoe het met de externe informatie moet omgaan en hoe het de output moet creëren.
- Externe informatie: is de extra informatie waar het LLM zijn antwoorden op kan baseren. Dit is de informatie die uit de vectordatabase wordt gehaald met RAG.
- Input gebruiker: de invoer van de gebruiker van het LLM.
- Output indicator: markeert het begin van de tekst die gegenereerd gaat worden (Briggs & Ingham, z.d.).

2.3 Beoordelen LLM en RAG

Om de output van het LLM, waarbij RAG is gebruikt, automatisch te beoordelen, is *retrieval augmented generation assessment* (RAGAs) ontwikkeld. RAGAs maakt gebruik van een LLM om de output te beoordelen, met als standaard GPT-3.5-turbo-16K. In totaal bestaat de RAGAs methode uit negen categorieën, waarvan er drie geen referentieantwoord vereisen. Het referentieantwoord, ook wel 'ground truth' genoemd, is het correcte antwoord op de vraag. In dit onderzoek is gekozen voor de volgende twee categorieën waar geen referentieantwoord voor nodig is en één categorie waar wel een referentieantwoord voor nodig is: faithfulness, answer relevance en answer correctness. Alle prestatiematen resulteren in een score tussen 0 en 1, waarbij een hogere score een betere prestatie aangeeft.

2.3.1 Faithfulness

Faithfulness is een prestatiemaat die evalueert of de output van het LLM feitelijk overeenkomt met de informatie uit de context. Deze evaluatiemethode is gekozen om de mate van hallucinaties te bepalen. De *faithfulness* score wordt bepaald door automatisch met een LLM een aantal claims te genereren op basis van het antwoord. Vervolgens worden deze claims gecontroleerd aan de hand van de context die gebruikt is om het antwoord te genereren. Dan wordt het aantal claims die waar zijn volgens de context gedeeld door het totaal aantal claims, zie Formule 1 (Es, James, Espinosa-Anke, & Schockaert, 2023).

$$\text{Faithfulness} = \frac{\text{Aantal claims die waar zijn volgens de context}}{\text{Totaal aantal claims}} \quad (1)$$

2.3.2 Answer relevance

Een andere prestatiemaat die gebruikt kan worden voor het beoordelen van RAG is *answer relevance*.

Answer relevance beoordeelt de relevantie van de output van het LLM ten opzichte van de input.

Deze prestatiemaat is gekozen om te bepalen of het model effectief antwoord geeft op de gestelde vraag. Deze methode houdt geen rekening met de waarheidsgrouwheid van het antwoord, maar straft antwoorden af die incompleet zijn of te veel herhalingen bevatten. De *answer relevance* score wordt bepaald door het volgende proces:

- Een vraag wordt aan het LLM gesteld en het LLM genereert een antwoord.
- Op basis van het gegenereerde antwoord, genereert een ander LLM een aantal potentiële vragen die passen bij dat antwoord.
- Voor elke potentiële vraag wordt de gelijkenis met de originele vraag berekend door middel van de cosinusgelijkenis. De cosinusgelijkenis geeft de gelijkenis tussen twee vectoren.
- Het gemiddelde van deze overeenkomsten wordt berekend, wat resulteert in de *answer relevance* score.

Zie formule 2 voor de formule van de *answer relevance* score.

$$\text{answer relevance score} = \frac{1}{n} \sum_{i=1}^n \text{sim}(q, q_i) \quad (2)$$

q = embedding van de originele vraag

q_i = embedding van de potentiële vraag

n = aantal potentiële vraag

Hiermee wordt beoordeeld hoe goed de output (het gegenereerde antwoord) aansluit bij de oorspronkelijke vraag, wat helpt om de relevantie en volledigheid van de antwoorden te meten (Es et al., 2023).

2.3.3 Answer correctness

De prestatiemaat *answer correctness* beoordeelt de correctheid van de antwoorden. In tegenstelling tot de twee eerdergenoemde prestatiematen, vereist deze maat wel een referentieantwoord. Bij deze prestatiemaat wordt er rekening gehouden met de betekenis van woorden en de feitelijke correctheid. *Answer correctness* wordt berekend door de som te nemen van de feitelijke correctheid en de semantische overeenkomst. De semantische overeenkomst is de overeenkomst van de

betekenis van woorden en zinnen. In Formule 3 is te zien hoe de feitelijke correctheid wordt berekend.

$$F_1 = \frac{TP}{(TP + 0,5 * (FP + FN))} \quad (3)$$

TP (True positive) =

Feiten die zowel in het referentieantwoord als in de output voorkomen

FP (False positive) =

Feiten die voorkomen in de output maar niet in het referentieantwoord

FN (False Negative) =

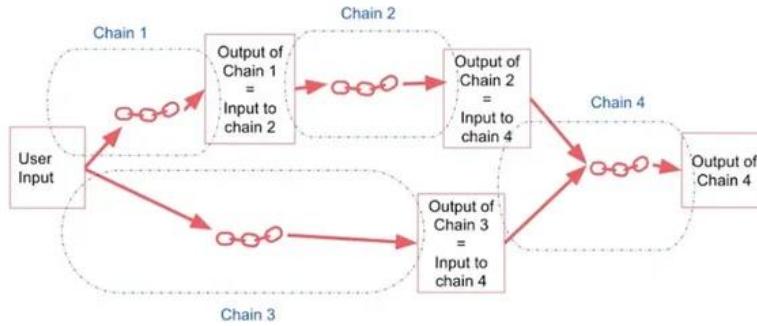
Feiten die voorkomen in het referentieantwoord maar niet in de output

Vervolgens wordt de semantische overeenkomst berekend met de cosinusgelijkenis. Dan wordt het gewogen gemiddelde berekend van de feitelijke overeenkomst en de semantische overeenkomst. Deze gewichten kunnen worden aangepast ('Answer Correctness', z.d.).

2.4 Chains en agents

Chains en agents zijn binnen Langchain twee manieren om een LLM een reeks acties uit te laten voeren. *Chains* bestaan uit een opeenvolging van *hardcoded* acties. Het model volgt dus altijd hetzelfde stappenplan. *Chains* kunnen worden onderverdeeld in drie categorieën: LLM-chain, sequential chain en router chain. De LLM-chain is de eenvoudigste *chain*, waarbij de prompt aan het LLM wordt gegeven en het LLM vervolgens antwoord geeft.

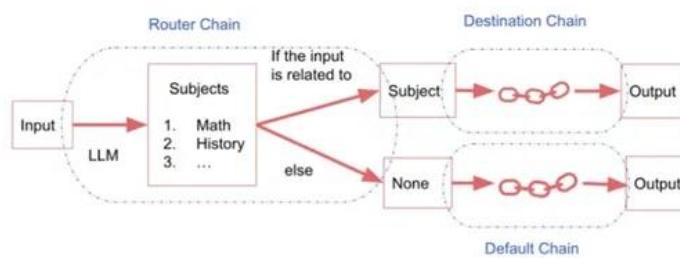
De *sequential chain* verbindt meerdere *chains* met elkaar, waardoor de output van de ene *chain* de input kan zijn van de andere *chain*. Zie Figuur 1 ter verduidelijking.



Figuur 1: schematische weergave van een sequential-chain (Banjara, 2023).

De *router chain* maakt met behulp van een LLM een keuze tussen meerdere *chains* en voert deze uit.

Zie Figuur 2.



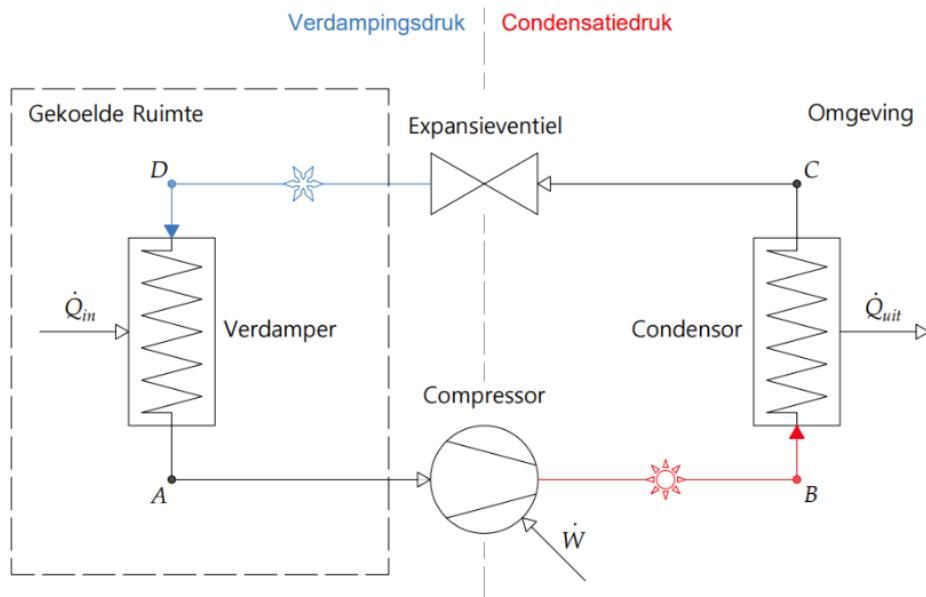
Figuur 2: schematische weergave van een router-chain (Banjara, 2023).

Agents stellen, net als *chains*, het LLM in staat om een reeks opvolgende acties uit te voeren. In tegenstelling tot *chains* zijn *agents* niet *hardcoded*. *Agents* hebben toegang tot verschillende tools, zoals een LLM of een *retriever* die informatie uit de vectordatabase haalt. Een LLM wordt gebruikt om te bepalen welke actie of tool moet worden uitgevoerd. De *agent* doorloopt daarbij de volgende stappen:

1. Bepaalt welke actie uitgevoerd moet worden op basis van de input van de gebruiker of de vorige output.
2. Doet de actie.
3. Observeert de output.
4. Herhaalt de eerste drie stappen tot dat het de taak heeft afgemaakt tot zijn beste vermogen (Smigel, 2023).

2.5 Koelsysteem

Om te bepalen welke onderdelen belangrijk zijn voor de prompt van het LLM, is het essentieel om de werking van een koelsysteem te begrijpen en bekend te zijn met de verschillende onderdelen. In Figuur 3 is een schematische weergave van een koelsysteem te zien.



Figuur 3: schematische weergave van een koelsysteem. Bron: FreezerData.

Een koelsysteem bestaat uit vier belangrijke elementen: de compressor, de condensor, het expansieventiel en de verdamper. Tussen deze elementen lopen leidingen met koelmiddel. Te beginnen bij de compressor, waar het koelmiddel uit de gekoelde ruimte binnen komt. De compressor verhoogt de druk. Aangezien een koelsysteem een gesloten systeem is, loopt de temperatuur van het koelmiddel op. De temperatuur wordt hoger dan de buitentemperatuur.

Vervolgens stroomt het koelmiddel naar de condensor, waar een deel van de warmte wordt afgegeven aan de buitenlucht. Dan komt het koelmiddel aan in het expansieventiel. Hier wordt de druk verlaagd waardoor ook de temperatuur lager wordt. Deze temperatuur is lager dan de temperatuur in de gekoelde ruimte. Daarna gaat het koelmiddel naar de verdamper, hier neemt het koelmiddel de warmte op uit de gekoelde ruimte, waardoor de gekoelde ruimte afkoelt. Ten slotte komt het koelmiddel terug in de compressor, waarna het hele proces zich herhaalt.

De sensoren van FreezerData zitten op drie punten in dit proces. De temperatuur van de heetgasleiding, vloeistofleiding en zuigleiding worden respectievelijk gemeten tussen de compressor en de condensor, tussen de condensor en het expansieventiel, en tussen de verdamper en de compressor.

3 Data

In paragraaf 3.1 wordt de data verkend die in het onderzoek gebruikt wordt. In paragraaf 3.2 wordt er ingegaan op de Fingerprint. De Fingerprint wordt gebruikt om afwijkend gedrag van het koelsysteem te identificeren.

3.1 Dataverkenning

In dit onderzoek wordt er gebruik gemaakt van data uit verschillende koelsystemen. In totaal is er toegang tot de gegevens van enkele honderden koelsystemen. De meetperiode varieert per systeem, maar omdat de data sinds augustus 2023 op een andere manier wordt opgeslagen en het inladen van gegevens van daarvoor erg lang duurt, is ervoor gekozen om alleen data vanaf augustus 2023 te gebruiken.

De gegevens worden ongeveer elke minuut verzameld met behulp van verschillende sensoren. Deze sensoren meten niet altijd op exact hetzelfde moment. Een voorbeeld hiervan is dat de temperatuur van de heetgasleiding op 08-02-2024 om 23:59:45 is gemeten en de temperatuur van de vloeistofleiding op 08-02-2024 om 23:59:46. Hierdoor ontbreekt er een meting van de vloeistofleiding om 23:59:45 en een meting van de heetgasleiding om 23:59:46.

Om dit probleem uniform op te lossen heeft FreezerData zijn eigen packages ontwikkeld. Hiermee worden onder andere de kolomnamen hernoemd en de frequentie van de waarnemingen aangepast, zodat per minuut één waarneming beschikbaar is. Wanneer er meerdere waarnemingen zijn binnen één minuut, wordt het gemiddelde van deze waarnemingen genomen. Indien er slechts één waarneming beschikbaar is in een minuut, wordt deze waarneming ingevoegd. Als er geen waarnemingen beschikbaar zijn in een minuut, wordt NaN ingevoerd.

Per koelsysteem verschilt het aantal sensoren waardoor ook het aantal variabelen per koelsysteem verschilt. Voor dit onderzoek worden de volgende variabelen gebruikt, deze variabelen zijn dus niet allemaal bij elk koelsysteem aanwezig:

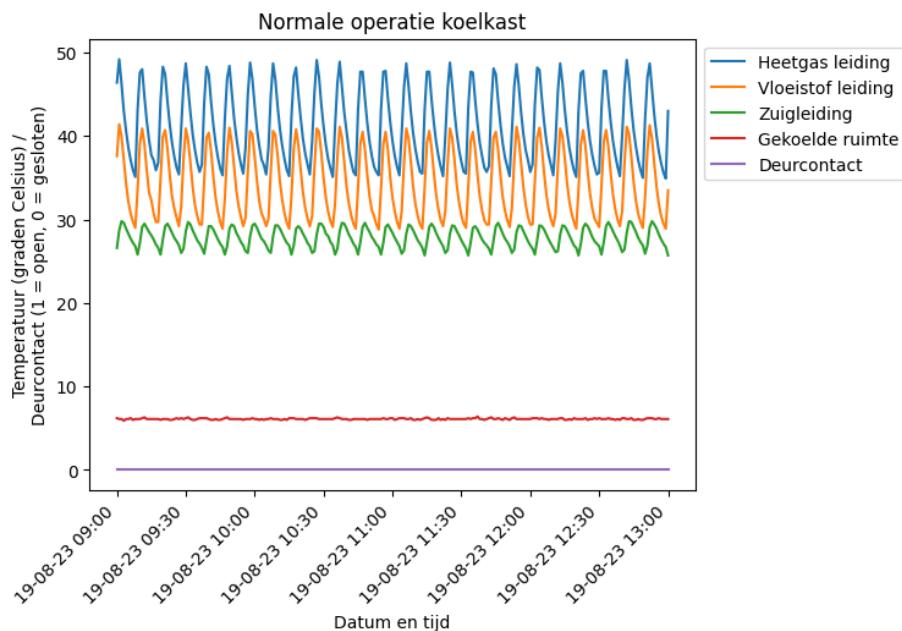
- Timestamp: geeft aan op welk moment de meting heeft plaatsgevonden.
- Deurcontact: geeft aan of de deur open is geweest tijdens de meting. 0 als de deur gesloten was, 1 als de deur open was.
- Gekoelde ruimte: geeft de temperatuur (graden Celsius) aan in de gekoelde ruimte.
- Heetgasleiding: geeft de temperatuur (graden Celsius) aan van de heetgasleiding.
- Vloeistofleiding: geeft de temperatuur (graden Celsius) aan van de vloeistofleiding.

- Zuigleiding: geeft de temperatuur (graden Celsius) aan van de zuigleiding.
- Omgeving: geeft de temperatuur (graden Celsius) aan van de omgeving.

3.1.1 Koelproces tijdens normale omstandigheden

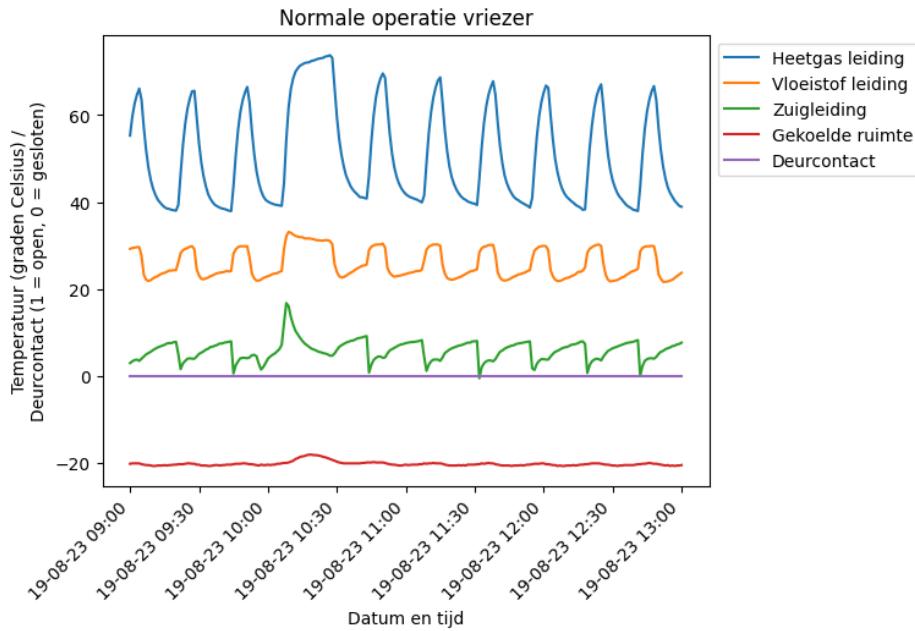
Om een beter beeld te krijgen bij de data worden hieronder twee figuren getoond waarbij een koelsysteem normaal functioneert.

In Figuur 4 is een deel van een tijdreeks te zien waar een koelkast normaal functioneert. Hierin valt op dat de koelkast niet constant aan staat, maar aan- en uitschakelt. Dit valt af te leiden aan de pieken van de temperatuur van alle leidingen. Daarnaast is de temperatuur van de gekoelde ruimte constant en ligt dit rond de zes graden Celsius.



Figuur 4: tijdreeks waar een koelkast normaal functioneert.

In Figuur 5 is een deel van een tijdreeks te zien waar een vriezer normaal functioneert. De temperatuur van de gekoelde ruimte is lager dan bij een koelkast en ligt rond de -20 °C. Omdat het koelsysteem meer moet koelen dan bij een koelkast wordt de temperatuur van de heetgasleiding hoger. Daarentegen zijn de temperaturen van de zuigleiding en vloeistofleiding juist lager dan bij een koelkast. Daarnaast is er rond 10:00 uur een afwijking in het patroon te zien. De temperaturen van de leidingen zijn hoger dan normaal en dit houdt langer aan dan gebruikelijk. Tegelijkertijd is de temperatuur van de gekoelde ruimte ook hoger dan normaal. Dit wordt veroorzaakt door het ontdooien van de verdamper, wat om de paar uur plaatsvindt. Voor de rest ziet het patroon van alle leidingen er over het algemeen hetzelfde uit als bij een koelkast.

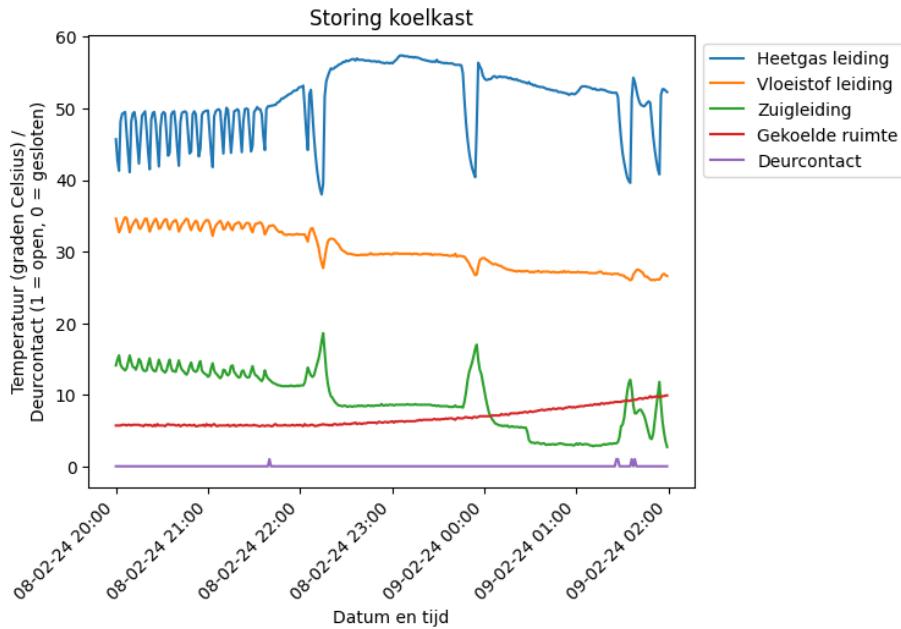


Figuur 5: tijdreeks waar een vriezer normaal functioneert.

De data varieert per koelsysteem, maar het algemene patroon lijkt bij elk koelsysteem overeen te komen. De variatie in de data heeft verschillende oorzaken, waaronder de mate van isolatie van het koelsysteem en het doel waarvoor het wordt gebruikt. Dit is bijvoorbeeld te zien in de afbeeldingen hierboven, waar het ene koelsysteem een koelkast is en het andere een vriezer.

3.1.2 Koelproces tijdens een storingen

In Figuur 6 is een periode van de tijdreeks te zien waar het hetzelfde koelsysteem als uit Figuur 4 een storing heeft. Hierin valt op dat het koelsysteem niet meer aan en uit schakelt. De temperatuur van de heetgasleiding wordt hoger dan normaal. De temperaturen van de vloeistofleiding en de zuigleiding worden juist lager dan normaal. De temperatuur van de gekoelde ruimte loopt op naar de omgevingstemperatuur.



Figuur 6: tijdreeks waar een koelkast een storing heeft.

3.2 Fingerprint

Om afwijkend gedrag van het koelsysteem te identificeren, maakt FreezerData gebruik van de Fingerprint. In de Fingerprint wordt per omgevingstemperatuur, met een maximum van 100 waarnemingen, het gemiddelde en de standaarddeviatie van de volgende variabelen bijgehouden: de maximale temperatuur per schakeling (ook wel piekwaarden genoemd) in de heetgasleiding, de vloeistofleiding en de zuigleiding, condensation pressure, evaporating pressure en run time. Een schakeling is het aangaan van de compressor.

Deze statistieken worden bijgehouden per omgevingstemperatuur, omdat het koelproces afhankelijk is van deze omgevingsfactor. Bij een hogere omgevingstemperatuur kan de condensor minder efficiënt warmte afvoeren aan de omgeving, waardoor de compressor harder moet werken en de temperaturen in de leidingen hoger worden.

Naast de omgevingstemperatuur wordt er in de fingerprint ook rekening gehouden met het deurcontact. Wanneer de deur van het koelsysteem open is geweest, worden de verzamelde gegevens van de komende dertig minuten niet toegevoegd aan de Fingerprint. Dit wordt gedaan omdat een open deur kan leiden tot een erg hoge piek in de temperatuur van de leidingen, wat niet representatief is voor de normale situatie wanneer het koelsysteem gesloten is, wat doorgaans het geval is gedurende het grootste deel van de tijd.

4 Methode

In paragraaf 4.1 wordt een automatische data-analyse ontwikkeld die de benodigde informatie uit de data haalt. In paragraaf 4.2 wordt de basis gelegd voor het RAG-framework. In paragraaf 4.3 wordt de opbouw van de prompt beschreven. In paragraaf 4.4 wordt de keuze van het LLM en de temperatuur uitgelegd.

4.1 Automatische data-analyse

Het LLM zal worden ingezet om monteurs te ondersteunen bij het repareren van koelsystemen. Het LLM kan de monteur advies geven op basis van informatie uit het koelsysteem en handleidingen. Aangezien een LLM waarschijnlijk niet in staat is de juiste informatie uit de ruwe data te halen, wordt er een automatische data-analyse ontwikkeld om het LLM toch belangrijke informatie te verschaffen over het koelsysteem. Deze analyse zal de belangrijkste bevindingen omzetten in tekst die wordt toegevoegd aan de prompt. Bij het ontwikkelen van de automatische data-analyse zal er rekening worden gehouden met de volgende punten:

- Het verschil tussen het gemiddelde van de maximale temperatuur per schakeling in de heetgasleiding van de afgelopen 24 uur en die van de normale periode uit de fingerprint.
- Het verschil tussen het gemiddelde van de maximale temperatuur per schakeling in de zuigleiding van de afgelopen 24 uur en die van de normale periode uit de fingerprint.
- Het verschil tussen het gemiddelde van de maximale temperatuur per schakeling in de vloeistofleiding van de afgelopen 24 uur en die van de normale periode uit de fingerprint.
- Het verschil tussen de gemiddelde run time van de compressor per schakeling van de afgelopen 24 uur en die van de normale periode uit de fingerprint.
- Het verschil tussen het gemiddeld aantal schakelingen per uur van de afgelopen 24 uur en die van de normale periode.
- De aanwezigheid van een trend in de temperatuur gedurende afgelopen 24 uur.
- Een abrupte verandering van de temperatuur gedurende afgelopen 24 uur
- Overschrijding van een maximale ingestelde waarde door de temperatuur in de gekoelde ruimte.
- Het aanstaan van de compressor in het afgelopen uur.
- De temperatuur van de heetgasleiding, zuigleiding, vloeistofleiding, gekoelde ruimte en omgeving.
- Het aantal minuten dat de deur open heeft gestaan gedurende de afgelopen 24 uur.

Aangezien FreezerData van plan is om het LLM bij alle koelsystemen in te zetten, is het van essentieel belang dat de data-analyse uniform en geschikt is voor elk koelsysteem. Daarom moet bij elk van de genoemde punten rekening worden gehouden met de variërende werking van elk koelsysteem.

4.1.1 Vergelijking van de gemiddelde maximale temperaturen

Het eerste deel van de automatische data-analyse richt zich op het vaststellen of de gemiddelde maximale temperaturen per schakeling van de heetgasleiding, vloeistofleiding en zuigleiding significant verschillen van de waarden in de fingerprint. Er is gekozen om de afgelopen 24 uur te vergelijken met de normale periode uit de fingerprint, omdat FreezerData hun algoritmes ook één keer in de 24 uur uitvoeren. Aangezien in de fingerprint de gemiddelde van de pieken wordt bijgehouden, worden in de afgelopen 24 uur eerst de piekwaarden van de verschillende leidingen bepaald met behulp van een package van FreezerData. Dit resulteert in een dataframe met de piekwaarden voor elke leiding.

Vervolgens wordt met een ongepaarde t-toets onderzocht of het gemiddelde van de piekwaarden van de afgelopen 24 uur significant verschilt van het gemiddelde van de piekwaarden in de fingerprint. Er is gebruik gemaakt van een ongepaarde t-toets omdat het gemiddelde van de afgelopen 24 uur en die uit de fingerprint afkomstig zijn uit twee verschillende groepen.

In bijna alle gevallen zijn er meer dan 30 waarnemingen beschikbaar, zowel in de fingerprint als in de dataframe met de piekwaarden. Als er meer dan 30 waarnemingen aanwezig zijn kan, vanwege de centrale limietstelling, worden aangenomen dat het steekproefgemiddelde bij benadering normaal verdeeld is. In uitzonderlijke gevallen wordt hier niet aan voldaan, zoals wanneer het koelsysteem net is geplaatst en de fingerprint onvoldoende is aangevuld. In deze gevallen is ervoor gekozen om de t-toets niet uit te voeren, waardoor het LLM over minder informatie zal beschikken. Er is bijvoorbeeld niet gekozen voor de Mann-Withney U toets, die in deze situatie wel gebruikt had kunnen worden, omdat in de fingerprint alleen de gemiddelde en standaarddeviatie beschikbaar zijn. Wat het uitvoeren van deze toets niet mogelijk maakt.

Bij het uitvoeren van de t-toets is er gekozen voor een relatief lage α -waarde van 0,001 om de kans op een fout van de eerste soort te minimaliseren. Met andere woorden, deze keuze vermindert de kans dat de t-toets een significant verschil in temperatuur aangeeft wanneer dat verschil in werkelijkheid niet aanwezig is. Het is van belang dat het LLM niet onterecht aangeeft dat er een probleem is met het koelsysteem. Als de t-toets wordt uitgevoerd, wordt het resultaat toegevoegd aan de prompt.

4.1.2 Vergelijking van de gemiddelde run time

In het tweede onderdeel van de data-analyse wordt er gekeken of de gemiddelde run time van de compressor in de afgelopen 24 uur significant verschilt van de gemiddelde run time uit de fingerprint. De run time is de tijd dat de compressor aanstaat per schakeling. Een schakeling is het aangaan van de compressor. Hierbij wordt opnieuw een ongepaarde t-toets gebruikt om te bepalen of het gemiddelde van de afgelopen 24 uur significant afwijkt van het gemiddelde uit de fingerprint. Er is bij deze variabele ook gekozen voor een ongepaarde t-toets omdat het gemiddelde van de afgelopen 24 uur en het gemiddelde uit de fingerprint afkomstig zijn van twee verschillende groepen. Ook deze twee steekproeven hebben bijna altijd meer dan 30 waarnemingen. Mocht er hier ook minder dan 30 waarnemingen aanwezig zijn, zal de t-toets ook niet uitgevoerd worden. Als de t-toets wel uitgevoerd kan worden zal dit ook gedaan worden met een α van 0,001 en wordt het resultaat toegevoegd aan de prompt.

4.1.3 Vergelijking van het gemiddeld aantal schakelingen

In het derde onderdeel van de automatische data-analyse wordt onderzocht of het gemiddelde van het aantal schakelingen per uur van de afgelopen 24 uur significant afwijkt van een normale periode. Aangezien deze variabele niet opgenomen is in de fingerprint van FreezerData wordt ervoor gekozen om voor de normale periode het gemiddelde van de voorgaande week te gebruiken. Deze benadering werkt effectief onder de veronderstelling dat de koelkast abrupt en niet geleidelijk defect raakt, dat een monteur binnen 24 uur na het optreden van een storing begint met repareren, en dat de koelkast niet binnen één week opnieuw defect raakt. Als aan één van deze voorwaarden niet wordt voldaan, kan deze benadering minder betrouwbaar zijn.

Voor zowel de periode van de afgelopen 24 uur en de periode van voorgaande week wordt het gemiddelde aantal schakelingen per uur bepaald door het gemiddelde aantal pieken per uur te berekenen. Om te toetsen of het gemiddeld aantal schakelingen per uur van de afgelopen 24 uur significant afwijkt van het gemiddelde uit de normale periode is er gekozen om de Mann-Whitney U toets te gebruiken, met een α van 0,001. Omdat het aantal waarnemingen van de afgelopen 24 uur altijd minder dan 30 is, vanwege de 24-urige tijdperiode, is de centrale limietstelling niet van toepassing. Bovendien kan niet voor elk koelsysteem gegarandeerd worden dat de steekproefverdeling normaal verdeeld is, waardoor de Mann-Whitney U toets een geschikte keuze is, aangezien deze toets niet vereist dat de steekproefverdeling normaal verdeeld is. Ook het resultaat van deze statistische toets wordt aan de prompt toegevoegd.

4.1.4 Detectie van trend in de temperatuur

In de automatische data-analyse wordt er bepaald of er in de afgelopen 24 uur een trend aanwezig is in de temperatuur van de heetgasleiding. Een opwaartse of neerwaartse trend kan impliceeren dat er een storing optreedt in het koelsysteem. Om te detecteren of er een trend aanwezig is, wordt de Mann-Kendall trend test gebruikt op de temperatuur van de heetgasleiding van de afgelopen 24 uur. De Mann-Kendall trend test is een non-parametrische toets, wat betekent dat de verdeling van de data niet bekend hoeft te zijn (Meals, Spooner, Dressing, & Harcum, 2011). Dit is van belang omdat in de automatische data-analyse niet altijd gegarandeerd kan worden dat de gegevens een bepaalde verdeling volgen. De nulhypothese van de Mann-Kendall trend test is dat de datapunten willekeurig zijn geordend en geen specifieke trend volgen. De alternatieve hypothese is dat de datapunten een monotone trend bevatten (Militino, Moradi, & Ugarte, 2020). Na het uitvoeren van de toets wordt het resultaat aan de prompt toegevoegd.

4.1.5 Detectie van verandering in de temperatuur

In de automatische data-analyse wordt gecontroleerd of er in de afgelopen 24 uur een verandering wordt gedetecteerd in de temperatuur van de heetgasleiding. Dit wordt gedaan met de Pettitt's test, een non-parametrische toets gebaseerd op de Mann-Whitney U toets. De nulhypothese van de test is dat de verdeling niet verandert in de tijd. De alternatieve hypothese is dat er een verandering is in de verdeling op een bepaald tijdstip (Militino et al., 2020). Nadat de Pettitt's test is uitgevoerd, wordt het resultaat hiervan toegevoegd aan de prompt.

4.1.6 Overige onderdelen automatische data-analyse

Naast de eerdergenoemde onderdelen van de automatische data-analyse, wordt er ook rekening gehouden met enkele andere onderdelen. Dit omvat de actuele temperatuur van de heetgasleiding, zuigleiding, vloeistofleiding, gekoelde ruimte en omgeving. Ook wordt het aantal minuten dat de deur in de afgelopen 24 uur open heeft gestaan aan de prompt toegevoegd, dit kan namelijk afwijkingen in de temperaturen verklaren. Daarnaast wordt door het analyseren van het aantal pieken in de heetgasleiding bepaald of de compressor het afgelopen uur heeft stilgestaan. Als er geen pieken zijn, heeft de compressor stilgestaan. Verder worden de algoritmes die FreezerData gebruikt aan de automatische data-analyse toegevoegd. Hierdoor kan worden gedetecteerd of er sprake is van een tekort aan koelmiddel, een bevroren verdamper en een verstopte condensor.

4.2 Implementatie van het RAG-framework

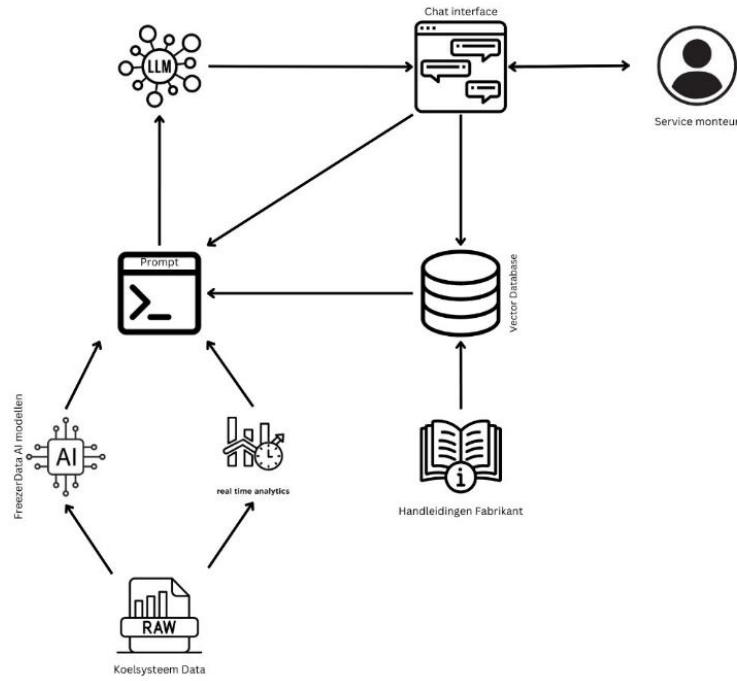
Om het LLM van domeinspecifieke kennis te voorzien, zoals handleidingen van koelkasten of vriezers, wordt de RAG-methode toegepast. Deze methode is beschreven in paragraaf 2.2. SURF, een coöperatieve vereniging van Nederlandse onderwijs- en onderzoeksinstellingen op het gebied van informatie- en communicatietechnologie, zou oorspronkelijk het RAG-framework leveren. Dit framework stond echter maximaal één pdf-document toe. Omdat het voor FreezerData essentieel is om meerdere PDF-documenten in een vectordatabase op te slaan, een koelsysteem kan meerdere handleidingen hebben, is zelf de basis van dit *framework* ontwikkeld. Hierdoor kunnen meerdere PDF-documenten worden ingeladen. De documenten worden opgesplitst in *chunks* van elk 1000 tokens, met een overlap van 200 tokens tussen de *chunks*. Dit zijn dezelfde instellingen die SURF hanteerde. Als vectordatabase is er gekozen voor Chroma DB en voor het embedding model is er gekozen voor het Text-embedding-ada-002-v2 van OpenAI. Vervolgens worden deze *chunks* opgeslagen in de vectordatabase. Bij het stellen van een vraag wordt de *chunk* die het meest overeenkomt met de vraag opgehaald uit de vectordatabase.

4.3 Proces en prompt

In paragraaf 4.3 wordt het proces beschreven waarmee het LLM informatie verkrijgt en de opbouw van de prompt besproken. De prompt is de input van het LLM. De informatie uit paragraaf 4.1 en 4.2 komt bij elkaar in de prompt.

4.3.1 Opbouw proces

In Figuur 7 wordt een schematische weergave van het proces getoond. De monteur stelt zijn vraag via de chatinterface, waarbij deze vraag direct aan de prompt wordt toegevoegd. Daarnaast wordt met deze vraag in de vectordatabase gezocht naar *chunks* die het meest overeenkomen, zoals beschreven in paragraaf 4.2. Verder wordt de data van het koelsysteem verwerkt door zowel de modellen van FreezerData als de automatische data-analyse, en dit wordt ook aan de prompt toegevoegd. Vervolgens wordt de prompt aan het LLM gegeven, die de vraag beantwoordt. De monteur kan dan het antwoord op zijn vraag zien in de chatinterface. Voor dit proces wordt een RetrieverQA-chain gebruikt als alternatief voor de LLM-chain die in paragraaf 2.4 wordt besproken. Bij deze chain wordt, naast de prompt die aan de LLM wordt gegeven, ook informatie uit de handleiding aan de prompt toegevoegd.



Figuur 7: schematische weergave van het proces.

4.3.2 Opbouw prompt

De prompt wordt opgebouwd zoals beschreven in paragraaf 2.2.2, waarbij wordt gestart met de instructies. Allereerst wordt beschreven wie het LLM is, wat zijn taken zijn, en aan wie zijn output gericht is. Daarnaast wordt vermeld dat het LLM gestraft wordt als het wiskundige termen gebruikt. Deze toevoeging is gedaan vanwege het feit dat de automatische data-analyse wiskundige termen bevat die de monteurs vermoedelijk niet kennen. Er is gekozen om te benadrukken dat het model straf krijgt voor het gebruik van wiskundige termen, in plaats van te verbieden dat deze termen gebruikt worden, vanwege de moeite die LLMs hebben met ontkenningen (Bsharat, Myrzakhan, & Shen, 2024).

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van IJskoud / FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Als je wiskundige termen gebruikt word je gestraft.

Om ervoor te zorgen dat het LLM alleen vragen beantwoordt met de informatie uit het koelsysteem of de handleidingen, zijn onderstaande regels aan de instructies toegevoegd. Het LLM moet eerst de informatie uit het koelsysteem gebruiken om de vraag te beantwoorden en vervolgens de informatie

uit de handleidingen raadplegen. Dit is toegevoegd om te voorkomen dat het LLM de verkeerde informatie gebruikt bij het formuleren van antwoorden.

Baseer je antwoord in eerste instantie op de informatie uit het koelsysteem.

*Als de informatie uit het koelsysteem niet voldoende is om de vraag te beantwoorden,
gebruik dan de context.*

Let op: gebruik de context alleen als die absoluut noodzakelijk is.

Als zelfs dat niet mogelijk is, geef dan aan dat je de vraag niet kunt beantwoorden.

Om de monteur een beter beeld te geven hoe het LLM tot een antwoord is gekomen, wordt de *reflection pattern* toegepast. De *reflection pattern* laat het LLM de redenering geven achter de output (White et al., 2023). Dit helpt de monteur bij het begrijpen van het antwoord en het achterhalen van mogelijke fouten.

Bij het geven van een antwoord, leg alsjeblieft de redenering en aanames achter je keuzen uit.

Gebruik specifieke voorbeelden of bewijs om je antwoord te ondersteunen waar mogelijk.

Behandel ook eventuele potentiele beperkingen in je antwoord om meer uitgebreide en nauwkeurige reactie te bieden.

Om het model beter te laten reageren, wordt er gevraagd of het model onbevoordeelde antwoorden genereert, en wordt het beloond met geld voor correcte antwoorden (Bsharat et al., 2024).

Zorg ervoor dat je antwoord onbevoordeeld is en vermijd het vertrouwen op stereotypen.

Ik geef je een beloning van \$300K voor een betere oplossing!

Naast de instructies worden ook de resultaten van de automatische data-analyse uit paragraaf 4.1 toegevoegd aan de prompt. Er is voor gekozen om deze informatie tussen '####' te plaatsen. Dit helpt het LLM om beter te begrijpen welke informatie bij het koelsysteem hoort (Bsharat et al., 2024). Dit gedeelte van de prompt ziet er als volgt uit:

Informatie uit het koelsysteem:

- De huidige temperatuur van de heet gasleiding is ... graden Celsius.*
- De huidige temperatuur van de vloeistofleiding is ... graden Celsius.*
- De huidige temperatuur van de zuigleiding is ... graden Celsius.*

- De huidige temperatuur van de gekoelde ruimte is ... graden Celsius.
- De huidige temperatuur van de omgeving is ... graden Celsius.
- Het koelsysteem heeft in de afgelopen 24 uur, ... minuten opengestaan.
- De temperatuur van de gekoelde ruimte is hoger dan de maximaal ingestelde temperatuur van 8 graden Celsius.
- De totale draaitijd van de compressor in de afgelopen 24 uur is ... uur.
- De gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur is ... minuten.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur en de gemiddelde draaitijd van de compressor per schakeling in een normale periode.
- Er is in de afgelopen 24 uur (g)een abrupte verandering in de temperatuur van de heetgasleiding gedetecteerd.
- Er is in de afgelopen 24 uur (g)een trend gevonden in de temperatuur van de heetgasleiding.
- Op basis van de statistische analyse is er (g)een significant verschil tussen het gemiddeld aantal schakelingen in de afgelopen 24 uur en het gemiddeld aantal schakelingen in een normale periode.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de heet gasleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de heet gasleiding in een normale periode.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de zuigleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de zuigleiding in een normale periode.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de vloeistofleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de vloeistofleiding in een normale periode.
- refrigerant shortage detected.
- Frozen evaporator detected
- Condensor clogging detected

###

De laatste drie regels van dit gedeelte van de prompt zijn het resultaat van de FreezerData-algoritmes. Deze regels worden alleen aan de prompt toegevoegd als de FreezerData-algoritmes een output hebben. Daarnaast wordt "(g)een" niet letterlijk aan de prompt toegevoegd. Afhankelijk van het resultaat van de uitgevoerde toets wordt er "geen" of "een" toegevoegd.

Vervolgens wordt de context met de informatie uit de handleidingen, de vraag van de gebruiker en de output indicator toegevoegd aan de prompt.

Context: {context}

Vraag: {query}

Antwoord:

Op de plaats {context} wordt de informatie uit de handleiding geplaatst die relevant is aan de vraag.

Op de plaats {query} wordt de vraag van de gebruiker ingevuld. De volledige prompt is te zien in

Bijlage A.

4.3.3 Problemen met proces

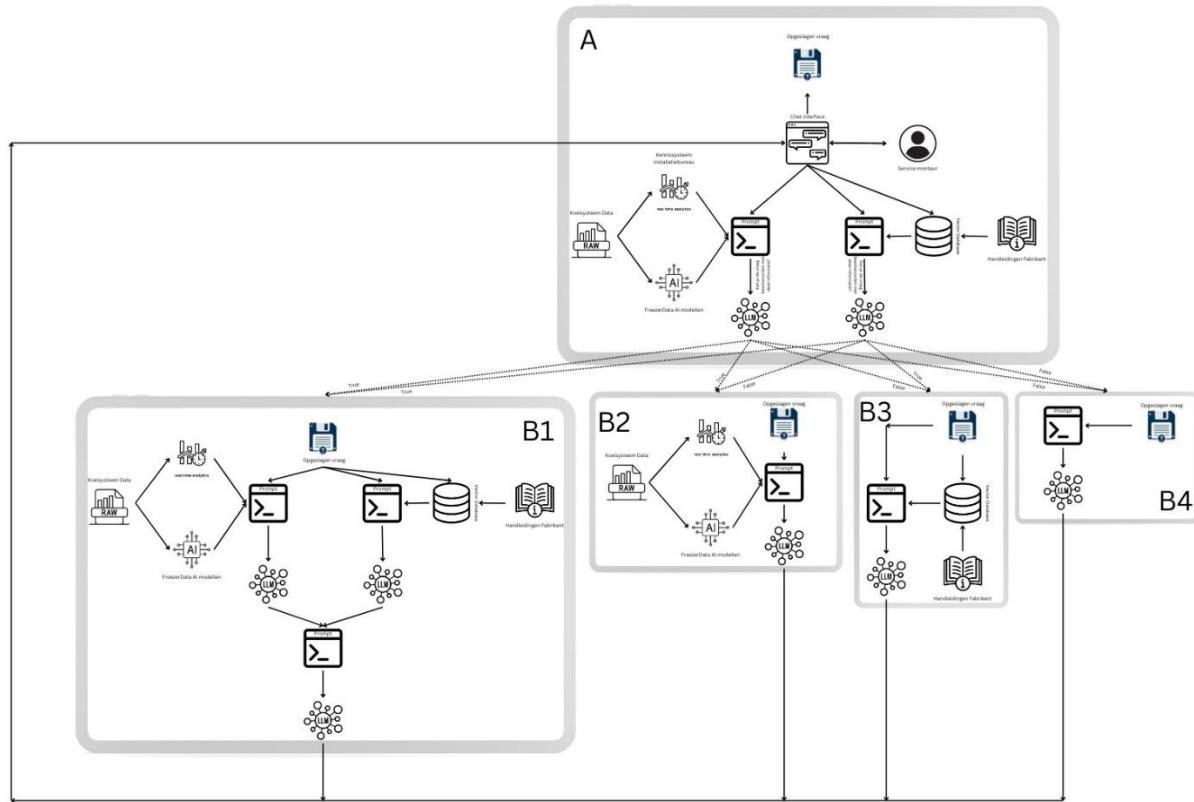
Bij deze manier van de prompt en het proces opbouwen doet zich soms het probleem voor dat het model moeite heeft om de juiste informatie te gebruiken bij het beantwoorden van vragen. Het LLM gebruikt bijvoorbeeld soms informatie uit de handleiding, terwijl dit niet relevant is voor het beantwoorden van de vraag. Neem bijvoorbeeld de vraag: "Moet er een monteur naar de koelkast komen?" Het LLM gaf het volgende antwoord:

"Ja, er moet een monteur naar de koelkast komen. De temperatuur van de zuigleiding is significant hoger dan normaal, wat kan duiden op een probleem met de koelinstallatie. Daarnaast is er een belemmering in de condensor gedetecteerd, wat kan leiden tot een verminderde prestatie van de koelinstallatie en een verhoogd energieverbruik. Aangezien de condensor regelmatig gereinigd moet worden en er tekenen van schade aan de compressor en condensor geconstateerd zijn, is het raadzaam dat een monteur de koelkast inspecteert en eventuele defecten repariert om optimale prestaties te garanderen."

In dit antwoord wordt onder andere aangegeven dat er tekenen van schade aan de compressor en condensor geconstateerd zijn. Deze informatie staat niet in de automatische data-analyse en heeft het model opgehaald in de vectordatabase. Deze informatie is niet juist voor de huidige staat van het koelsysteem, maar het LLM presenteert het alsof het wel klopt. Hoewel er pogingen zijn gedaan om dit probleem op te lossen door middel van aanpassingen aan de prompt, was dit het beste resultaat dat is bereikt.

4.3.4 Verbetering proces en prompt

Om dit probleem toch op te lossen is het proces waarmee informatie aan het LLM wordt verstrekt gewijzigd. In Figuur 8 is een schematische weergave te zien (Zie Bijlage B voor vergrote weergave).



Figuur 8: schematische weergave van het verbeterde proces.

4.3.4.1 Prompt actie A

In het nieuwe proces stelt de monteur nog steeds zijn vraag via de chatinterface. Vervolgens wordt deze vraag toegevoegd aan twee verschillende prompts. Dit gebeurt in actie A (zie Bijlage B, Figuur 1B voor een schematische weergave). Eén prompt bevat de resultaten van de automatische data-analyse, en de andere prompt bevat informatie uit de handleidingen die relevant zijn voor de gestelde vraag. Dit wordt op dezelfde manier toegevoegd als in de voorgaande methode. In beide prompts wordt de vraag gesteld of de verstrekte informatie voldoende is om de vraag te beantwoorden met als output True of False. Met de True of False output wordt vervolgens de keuze gemaakt welke prompt hierna gebruikt wordt. Normaal zou het gebruik van een router chain voor de hand liggen. Echter werkt de router chain niet in combinatie met de retrieverQA chain die nodig is om de relevante informatie uit de handleidingen aan de prompt toe te voegen. Daarom is er gekozen voor deze methode. De instructies in de prompt zien er als volgt uit:

Prompt 1

Instructies: Beoordeel of je de vraag kan beantwoorden met uitsluitend de informatie uit het koelsysteem. Je antwoord moet True of False zijn.

Prompt 2

Instructies: Beoordeel of je de vraag kan beantwoorden met uitsluitend de context uit de handleidingen. Je antwoord moet True of False zijn.

Op basis van het resultaat van actie A wordt één van de volgende acties uitgevoerd:

- Als de output van beide prompts True is, wordt Actie B1 uitgevoerd. Hierbij wordt de vraag beantwoord op basis van zowel de data-analyse als de handleidingen.
- Als alleen de automatische data-analyse voldoende is, wordt Actie B2 uitgevoerd. Hierbij wordt de vraag beantwoord op basis van de data-analyse.
- Als alleen de informatie uit de handleidingen voldoende is, wordt Actie B3 uitgevoerd. Hierbij wordt de vraag beantwoord met de informatie uit de handleidingen.
- Als zowel de automatische data-analyse als de informatie uit de handleidingen niet voldoende zijn, wordt Actie B4 uitgevoerd. Hierbij wordt er geen antwoord gegeven op de vraag.

Een schematische weergave van deze acties staat in Bijlage B.

4.3.4.2 Prompt actie B1

In actie B1 wordt de vraag apart beantwoord met de informatie uit de automatische data-analyse (zie prompt 1) en apart beantwoord met de informatie uit de handleidingen (zie prompt 2) (zie Bijlage B, Figuur 2B voor een schematische weergave). Actie B1 wordt uitgevoerd met een sequential-chain zoals besproken in paragraaf 2.4. De instructies van deze prompts zien er als volgt uit:

Prompt 1

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Als je wiskundige termen gebruikt word je gestraft.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit het koelsysteem. Als dit niet lukt geef je aan dat je de vraag niet kan beantwoorden.

Prompt 2

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit de context. Als dit niet lukt geef je aan dat je de vraag niet kan beantwoorden.

Vervolgens worden de resultaten van deze twee prompts samengevoegd in een nieuwe prompt. Het LLM beoordeelt met behulp van deze prompt welk van de twee antwoorden het meest geschikt is en geeft dat antwoord aan de monteur. De prompt ziet er als volgt uit:

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

De monteurs zijn experts op het gebied van koelsystemen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Jouw taak is om te beoordelen welk antwoord op de vraag het beste is.

Vraag: {query}

Antwoord gebaseerd op een data – analyse van het koelsysteem: {antwoord_1}

Antwoord gebaseerd op de handleiding van het koelsysteem: {antwoord_2}

Je output is het antwoord dat jij als beste hebt beoordeeld. Dus alleen antwoord gebaseerd op een data – analyse van het koelsysteem of alleen antwoord gebaseerd op de handleiding van het koelsysteem. Je wordt gestraft als je output anders is dan het beste antwoord.

4.3.4.3 Prompt actie B2

In actie B2 wordt de vraag beantwoord met de informatie uit de automatische data-analyse, waarna het antwoord aan de monteur wordt verstrekt (zie Bijlage B, Figuur 3B voor een schematische weergave). Deze actie wordt uitgevoerd met een LLM-chain. De instructies in de prompt zien er als volgt uit:

Instructies: Je bent Kelvin Fahrenheit, de virtuele monteur van IJskoud / FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn expert op het gebied van koelsystemen.

Als je wiskundige termen gebruikt wordt je gestraft.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit het koelsysteem. Als dit niet lukt geef je aan dat je de vraag niet kan beantwoorden.

4.3.4.4 Prompt actie B3

In actie B3 wordt de vraag beantwoord met de informatie uit de handleidingen. Vervolgens wordt dit antwoord aan de monteur verstrekt (zie Bijlage B, Figuur 4B voor een schematische weergave). Deze actie wordt uitgevoerd met de *retrieverQA-chain*. De instructies in de prompt zien er als volgt uit:

Instructies: Je bent Kelvin Fahrenheit, de virtuele monteur van IJskoud / FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn expert op het gebied van koelsystemen.

Als je wiskundige termen gebruikt wordt je gestraft.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit de context. Als dit niet lukt geef je aan dat je de vraag niet kan beantwoorden.

4.3.4.5 Prompt actie B4

In actie B4 wordt aangegeven dat de vraag niet beantwoord kan worden (zie Bijlage B, Figuur 5B voor een schematische weergave). Deze actie wordt uitgevoerd met een *LLM-chain*. De instructie in de prompt ziet er als volgt uit:

Instructies: Jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData. Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud. Twee andere virtuele monteurs hebben bepaald dat deze vraag niet beantwoord kan worden. Geef netjes aan dat je deze vraag niet kan beantwoorden.

Alle volledige prompts zijn te zien in Bijlage A.

4.3.5 Afwegingen nieuw proces

Dit nieuwe proces heeft als voordeel dat er geen informatie uit de handleidingen kan worden toegevoegd die niet overeenkomt met de huidige staat van het koelsysteem. Dit zorgt ervoor dat de output van het LLM minder verkeerde informatie bevat. Daarnaast biedt deze oplossing de mogelijkheid om de prompt per onderdeel aan te passen, zodat de instructies beter kunnen worden afgestemd op de data-analyse of de informatie uit de handleidingen. Deze oplossing brengt echter ook een nadeel met zich mee. De informatie uit de data-analyse en de handleidingen zal nooit gecombineerd worden bij het genereren van een antwoord. Het antwoord zal dus altijd uitsluitend gebaseerd zijn op ofwel de data-analyse, ofwel de informatie uit de handleidingen.

4.4 Large Language model

De prompts uit paragraaf 4.3 zijn de input van het LLM. De keuze van het LLM is belangrijk en heeft naast de prompts invloed op de gegenereerde output. In overleg met FreezerData is er gekozen voor een LLM van OpenAI, wat hun voorkeur had. Uiteindelijk is er gekozen voor GPT-3.5 Turbo. Dit model is 20 keer goedkoper dan het beste model van OpenAI, GPT-4 Turbo. Daarnaast is GPT-3.5 Turbo aanzienlijk sneller dan GPT-4 Turbo. De combinatie van de lagere kosten en de hogere snelheid heeft geleid tot de keuze voor GPT-3.5 Turbo, ondanks het feit dat GPT-4 Turbo een beter model is.

Naast de keuze van het LLM is ook de temperatuurinstelling belangrijk. De temperatuur is een hyperparameter die willekeur van de output beïnvloedt. De waarde van de temperatuur varieert tussen de nul en twee. Een lage temperatuur zorgt voor een consistente output, een hoge temperatuur zorgt voor een willekeurigere output ('OpenAI Platform', z.d.). Om hallucinaties bij het model te voorkomen is de temperatuur van het model aangepast. Bij de prompts waar de output beperkt is tot alleen True of False en bij de prompt waarbij de output een keuze is tussen twee antwoorden, is een temperatuur van 0 gekozen. Deze lage temperatuur minimaliseert de mate van willekeur in het antwoord ('LLM Settings', 2024), waardoor het model eerder geneigd is om zich aan de gevraagde output in de prompt te houden. Bij de prompt waarbij er een antwoord wordt gegenereerd op de vraag van de monteur, is er gekozen voor een temperatuur van 0,3. Dit geeft het model meer vrijheid in de keuze van woorden, waardoor het model niet elke keer exact hetzelfde antwoord genereert. Door de relatief lage temperatuur blijft het risico op hallucinaties echter nog steeds laag.

5 Evaluieren en analyseren output LLM

In dit hoofdstuk staat de evaluatie en analyse van de output van het LLM centraal. Het is van essentieel belang om de prestaties van het LLM te beoordelen om de bruikbaarheid en betrouwbaarheid ervan te kunnen inschatten. Om de output van het LLM te beoordelen, worden de prestatiematen uit paragraaf 2.3 gebruikt. Hierbij worden de mate van hallucinatie, de relevantie van het antwoord en de correctheid van het antwoord geëvalueerd. Voor de beoordeling van de output van het LLM worden er een aantal vragen opgesteld die een monteur/gebruiker potentieel zou kunnen stellen. Deze vragen zijn deels bedacht met FreezerData en deels bedacht om alle elementen van het proces te beoordelen. De vragen zijn onderverdeeld in drie categorieën: vragen die beantwoord moeten worden op basis van de data, vragen die beantwoord moeten worden op basis van de handleidingen en vragen die het LLM niet mag beantwoorden.

Vragen die beantwoord moeten worden op basis van de data

- Hoe gaat het met de koelkast?
- Moet er een monteur naar de koelkast komen?
- Wat zijn alle huidige temperaturen?

Vragen die beantwoord moeten worden op basis van de handleidingen

- Ik zie foutcode F7, wat moet ik doen?
- Hoe stel ik een lokaal deuralarm in?

Vragen die het model niet moet beantwoorden

- Waar moet ik naartoe op vakantie?
- Wat ligt er in de koelkast opgeslagen?

Bij de vragen die op basis van de data beantwoord moeten worden, is het antwoord afhankelijk van de data. Om dit te beoordelen, wordt dezelfde vraag gesteld in zes situaties met een storing en in twee situaties waarin het koelsysteem normaal functioneert. Er is gekozen om de output van het model vaker te beoordelen in situaties met een storing dan in situaties waarin de koelkast normaal functioneert, omdat er eerder vragen worden gesteld wanneer er zich een probleem voordoet. De monteur stelt namelijk vragen wanneer hij bij de koelkast aanwezig is, en meestal is dat het geval wanneer er een storing is. Vervolgens wordt het gemiddelde van de scores per vraag berekend. De output van elke vraag en de bijbehorende scores zijn te vinden in Bijlage C.

In Tabel 1 worden de gemiddelde scores weergegeven voor elke vraag, die moet worden beantwoord op basis van de data. De scores omvatten *faithfulness*, *Answer relevance* en *Answer correctness*.

Tabel 1: de gestelde vragen aan het model die beantwoord moeten worden op basis van de data met de bijbehorende scores.

Vraag	Faithfulness	Answer relevance	Answer correctness
Hoe gaat het met de koelkast?	0,94375	0,825679	0,550728
Moet er een monteur naar de koelkast komen?	0,979166	0,827490	0,720723
Wat zijn alle huidige temperaturen?	1	0,872345	0,97719

In Tabel 2 worden de scores weergegeven voor elke vraag, die moet worden beantwoord op basis van de handleidingen. De scores omvatten *faithfulness*, *Answer relevance* en *Answer correctness*.

Tabel 2: de gestelde vragen aan het model die beantwoord moeten worden op basis van de handleidingen met de bijbehorende scores.

Vraag	Faithfulness	Answer relevance	Answer correctness
Ik zie foutcode F7, wat moet ik doen?	1	0,874554	0,867115
Hoe stel ik een lokaal deuralarm in?	1	0,963394	0,766729

Bij de vragen die op basis van de handleidingen beantwoord moeten worden, behaalt het LLM een *faithfulness* score van 1, wat betekent dat het antwoord uitsluitend informatie bevat die in de context is meegegeven. Bij de vragen die gebaseerd zijn op data, is deze score gelijk of lager. Hierin bevat de output van het model soms informatie die niet in de data-analyse staat. Hoewel de informatie niet altijd onjuist is, brengt dit wel een risico op hallucinaties met zich mee. Een verklaring voor deze lagere score is dat het antwoord op deze vraag niet letterlijk in de data-analyse staat, maar eerder een conclusie is die het model trekt op basis van de informatie in de data-analyse. Hierdoor kan het voorkomen dat bij het berekenen van de *faithfulness* score een claim gegenereerd wordt die in de output van het model zit maar niet voorkomt in de data-analyse (de meegegeven context),

bijvoorbeeld: 'Het gaat goed met de koelkast'. Hoewel dit een correct antwoord zou kunnen zijn, kan dit leiden tot een lagere *faithfulness* score.

De *answer relevance* score is bij de vragen op basis van data iets lager dan bij de vragen op basis van de handleidingen. Deze scores variëren van ongeveer 0,8 tot 0,95. Dit betekent dat het antwoord relevant is op de gestelde vraag, maar soms onvolledig is of overbodige informatie bevat.

De *answer correctness* score is bij alle categorieën relatief laag, behalve bij de vraag "Wat zijn alle huidige temperaturen?". Bij de vragen die op basis van de handleidingen worden beantwoord, komt dit doordat het model soms informatie weglaat die wel in de handleidingen staat. De informatie uit de handleidingen is exact overgenomen in de ground truth. In sommige gevallen is deze informatie essentieel, in andere gevallen is de missende informatie niet essentieel, maar beïnvloedt het de score wel negatief. Daarnaast worden sommige tabellen en symbolen uit de handleidingen niet goed uitgelezen. De symbolen in de handleidingen zijn geen tekst en worden hierdoor niet opgehaald. Hierdoor bevat de output van het LLM geen symbolen. Dit is niet terug te zien in de score omdat deze symbolen ook niet in de ground truth vermeld kunnen worden, maar het heeft wel invloed op de bruikbaarheid van het antwoord.

Bij vragen die op basis van data worden beantwoord, is de *answer correctness* score lager omdat de antwoorden van het model soms informatie bevatten uit de automatische data-analyse die niet in de ground truth staat. Dit leidt tot een lagere score, zelfs als de informatie niet onjuist is. De uitzondering is de vraag "Wat zijn alle huidige temperaturen?". Voor deze vraag is de *answer correctness* score relatief hoog, omdat de huidige temperaturen direct uit de data-analyse kunnen worden afgeleid zonder extra interpretatie of aanvulling. Dit betekent dat er minder snel een verschil zit tussen het antwoord van het model en de ground truth. Daarnaast kan soms de *answer relevance* score bij bepaalde vragen misleidend zijn vanwege de algoritmes van FreezerData. Deze algoritmes geven soms aan dat er een bevroren verdamper is, terwijl dit in werkelijkheid niet het geval is. Als dit niet daadwerkelijk is gebeurd, wordt het ook niet vermeld in de 'ground truth', wat resulteert in een lagere score voor de *answer relevance*.

In Tabel 3 staan de vragen die binnen de categorie vallen die het model niet mag beantwoorden. Er zijn hierbij geen scores gebruikt, omdat deze niet relevant zijn.

Tabel 3: vragen die het model niet mag beantwoorden met het bijbehorende antwoord.

Vraag	Antwoord
Waar moet ik naartoe op vakantie?	Sorry, maar ik kan je niet helpen met deze vraag.
Wat ligt er in de koelkast opgeslagen?	Sorry, maar ik kan deze vraag niet beantwoorden.

De eerste vraag heeft geen betrekking op koelsystemen. Het model geeft aan dat het deze vraag niet kan beantwoorden. De tweede vraag gaat wel over koelsystemen, maar het model zou niet in staat kunnen zijn om deze vraag correct te beantwoorden met de verkregen informatie. Ook hier geeft het model aan dat het de vraag niet kan beantwoorden.

Naast de RAGAs-methode zijn de antwoorden van het LLM beoordeeld door een expert in koelsystemen. De resultaten van deze beoordeling zijn te zien in Tabel 4. Elke vraag is beoordeeld met een van de volgende waarderingen: onvoldoende, matig, voldoende, ruim voldoende of goed.

Tabel 4: beoordeling van de output van het LLM door de expert per vraag.

Vraag	Beoordeling						Totaal
	Onvoldoende	Matig	Voldoende	Ruim voldoende	Goed		
Hoe gaat het met de koelkast?	2	0	0	3	3		8
Moet er een monteur naar de koelkast komen?	2	0	5	1	0		8
Wat zijn alle huidige temperaturen?	0	0	0	0	8		8
Ik zie foutcode F7, wat moet ik doen?	0	0	0	0	1		1
Hoe stel ik een lokaal deuralarm in?	0	0	0	1	0		1

In de tabel is te zien dat bijna alle antwoorden beoordeeld worden met een voldoende of hoger. Bij de vragen "Hoe gaat het met de koelkast?" en "Moet er een monteur naar de koelkast komen?" zijn twee antwoorden als onvoldoende beoordeeld. Het LLM gaf hierbij aan dat er een bevoren verdamper aanwezig was, wat niet het geval was. Dit komt, zoals eerder uitgelegd, door de algoritmes van FreezerData. Over het algemeen beoordeelde de expert de antwoorden van het LLM met een voldoende of hoger en bevatten ze de juiste conclusies. Soms werden de antwoorden echter als algemeen beschouwd, vooral wanneer de algoritmes van FreezerData geen storing herkenden.

6 Conclusie

In dit rapport is onderzocht hoe een *Large Language Model* (LLM), op basis van koelsysteemdata en handleidingen, vragen van monteurs kan beantwoorden. Dit onderzoek omvatte de volgende deelopdrachten:

- Het ontwikkelen van een automatische data-analyse.
- Het ontwikkelen van een basis Retrieval-Augmented generation (RAG)-framework.
- Het opstellen van een prompt.
- Het ontwikkelen van een proces dat bepaalt welke informatie in de prompt wordt verwerkt.
- Het evalueren van de output van het LLM.

Het doel van het inzetten van een LLM is om de benodigde tijd van de werkzaamheden van de monteurs te verkorten, wat kan leiden tot een vermindering van het personeeltekort en de werkdruk.

Door het ontwikkelen van een automatische data-analyse en door gebruik te maken van *Retrieval-Augmented Generation* (RAG) is het mogelijk om een LLM te ontwikkelen dat vragen van monteurs kan beantwoorden op basis van koelsysteemdata en handleidingen. De ontwikkelde automatische data-analyse haalt de essentiële informatie uit de koelsysteemdata, terwijl het RAG-framework de relevante informatie ophaalt uit de handleidingen. Er is een proces ontwikkeld die bepaalt welke informatie verwerkt wordt in de opgestelde prompt. Met de opgestelde prompt kan het LLM output genereren.

De evaluatie van de output van het LLM toont aan dat de antwoorden niet tot nauwelijks hallucinaties bevatten, wat blijkt uit de hoge *faithfulness* scores. Daarnaast blijkt uit de *answer relevance* score dat het LLM over het algemeen relevante antwoorden gaf, hoewel sommige antwoorden soms onvolledig waren of overbodige informatie bevatten. De *answer correctness* werd lager beoordeeld, maar de expert beoordeelde de antwoorden over het algemeen als voldoende met de juiste conclusies. De beoordeling van de expert suggereert dat de implementatie van het LLM de potentie heeft om de werkdruk van de monteurs te verlagen door snelle en nauwkeurige antwoorden te genereren op de vragen van de monteurs. Echter, zaten er nog enkele beperkingen aan de antwoorden van het LLM. In sommige gevallen, vooral wanneer de algoritmes van FreezerData geen output hadden, werden de antwoorden als te algemeen en niet specifiek genoeg beschouwd.

Ondanks enkele uitdagingen kan het inzetten van een LLM een veelbelovende oplossing zijn voor het verminderen van werkdruk van monteurs bij Ijskoud. Het gebruik van een LLM kan de werkzaamheden van monteurs ondersteunen en versnellen. Dit onderzoek is een sterke basis voor het doorontwikkelen van dit systeem.

7 Discussie

Het doel van dit rapport was om te onderzoeken hoe een Large Language model (LLM) op basis van koelsysteemdata en handleidingen monteurs kan ondersteunen bij hun werkzaamheden. Door de inzet van het LLM hoopt IJskoud de werkdruk onder de monteurs te verminderen.

Ondanks de veelbelovende resultaten zijn er enkele beperkingen. In dit onderzoek werden acht situaties beoordeeld, waarvan zes een storing waren en twee tijdens normale werking van het koelsysteem. In deze situaties beoordeelde het LLM altijd correct of er een probleem aanwezig was. Er kan echter niet gegarandeerd worden dat het LLM altijd juist beoordeelt of er een storing aanwezig is. Bovendien moet de storing ook terug te zien zijn in de data. Als dit niet het geval is, kan het LLM niet weten dat er een storing aanwezig is.

Verder is het model beoordeeld op een beperkt aantal vragen. De evaluatie laat zien dat het antwoord op deze vragen soms beperkingen had, maar over het algemeen goed beantwoord werd. Echter, kan er niet gegarandeerd worden dat andere vragen ook juist beantwoord worden.

Deze punten laten zien dat vervolgonderzoek naar een uitgebreidere evaluatie van het LLM in verschillende situaties nuttig kan zijn. Het testen van het LLM op verschillende soorten storingen, normale omstandigheden en blootstellen aan meerdere vragen kan helpen bij het identificeren van eventuele zwakke en verbeterpunten, maar ook de sterke punten van het LLM naar voren brengen. Het betrekken van monteurs bij het test- en feedbackproces kan ook waardevolle inzichten opleveren en de acceptatie van het LLM in de praktijk inzetten bevorderen.

De opbouw van het huidige proces brengt ook een beperking met zich mee. De informatie uit de data-analyse kan niet gecombineerd worden met de informatie uit de handleidingen. Hierdoor kunnen de antwoorden alleen gebaseerd zijn op óf de data-analyse óf de handleidingen, maar niet op een combinatie van beide. Dit zou ook één van de oorzaken kunnen zijn dat de antwoorden soms als te algemeen werden beschouwd. Er kan namelijk nooit met de bevindingen uit de data-analyse gezocht worden naar mogelijke problemen in de handleidingen. Er is op twee manieren geprobeerd de informatie uit de data-analyse te combineren met de informatie uit de handleidingen. Er is geprobeerd om het huidige proces uit te breiden (zie bijlage D) en er is geprobeerd om gebruik te maken van agents. Dit leverde allebei niet het gewenste resultaat op.

Daarnaast zijn er nog beperkingen in het huidige RAG-framework. Momenteel worden alle handleidingen die in de vectordatabase staan opgeslagen gebruikt om antwoord te genereren, in plaats van alleen de handleidingen van het koelsysteem waar op dat moment een vraag over wordt gesteld. Bovendien worden sommige tabellen en symbolen niet correct uitgelezen. Dit basis framework wordt verder ontwikkeld door een student van de Universiteit van Amsterdam (UvA). Hij zal ervoor zorgen dat alleen handleidingen van de koelkast worden gebruikt waar op dat moment een vraag over wordt gesteld. Daarnaast zal hij ervoor zorgen dat tabellen en symbolen correct worden uitgelezen, wat momenteel niet altijd gebeurt. Het vervolgonderzoek van de UvA-student is van belang voor een optimale werking van dit framework.

8 Aanbevelingen

Aan de hand van de bevindingen uit de voorgaande hoofdstukken wordt het volgende advies aan FreezerData gegeven:

- Er wordt aanbevolen om het LLM in eerste instantie in productie te brengen, waarbij gebruikers alleen de mogelijkheid hebben om een aantal vooraf bepaalde vragen te stellen. Er kan immers niet gegarandeerd worden dat het LLM correcte antwoorden zal geven op andere vragen. Als er in de toekomst meer zekerheid is dat het model ook correct antwoord geeft op andere vragen, kan overwogen worden om er een chatbot van te maken.
- Er wordt aanbevolen om de output van het LLM uitgebreider te evalueren en hier indien nodig monteurs bij te betrekken. Dit kan helpen bij een beter identificatie van de mogelijke zwakke en verbeterpunten.
- Momenteel worden vragen door het model uitsluitend beantwoord op basis van de data-analyse of de handleidingen. Er wordt aanbevolen het proces te verbeteren, zodat het LLM de informatie uit de data-analyse en de handleidingen kan combineren. Er is op twee manieren geprobeerd dit proces te verbeteren: het uitbreiden van het huidige proces (zie bijlage D) en het gebruik van *agents*. Hoewel beide oplossingen nog niet de gewenste resultaten opleverden, hebben ze potentie. Het wordt aanbevolen om één van de oplossingen verder te ontwikkelen.
- Er wordt aanbevolen om de automatische data-analyse in de toekomst uit te breiden met nieuwe algoritmes en AI-modellen van FreezerData, zodat het LLM nog meer informatie over het koelsysteem heeft.
- Er wordt aangeraden om het RAG-framework van de UvA-student te gebruiken, zodat het LLM uitsluitend de handleidingen van de betreffende koelkast raadpleegt bij het beantwoorden van vragen. Daarnaast worden tabellen en symbolen beter uitgelezen.
- Er wordt aanbevolen om geheugen aan het LLM toe te voegen, met name als er in de toekomst een chatbot van het LLM gemaakt wordt. Het toevoegen van geheugen zal waarschijnlijk resulteren in vloeiendere gesprekken.
- Vanwege de snelle ontwikkelingen op het gebied van LLM, wordt er aanbevolen om in de toekomst mogelijk voor een ander LLM te kiezen. Houd er echter rekening mee dat een ander LLM anders kan reageren op dezelfde prompt.

Literatuurlijst

Aanpak personeelstekort in techniek en ICT [Nieuwsbericht]. (2023, februari 3). Geraadpleegd 20 mei 2024, van Rijksoverheid website:

<https://www.rijksoverheid.nl/actueel/nieuws/2023/02/03/aanpak-personeelstekort-in-techniek-en-ict>

Answer Correctness. (z.d.). Geraadpleegd 15 maart 2024, van Ragas website:

https://docs.ragas.io/en/stable/concepts/metrics/answer_correctness.html

Banjara, B. (2023, oktober 19). A Comprehensive Guide to Using Chains in Langchain. Geraadpleegd 28 mei 2024, van Analytics Vidhya website:

<https://www.analyticsvidhya.com/blog/2023/10/a-comprehensive-guide-to-using-chains-in-langchain/>

Barnett, S., Kurniawan, S., Thudumu, S., Brannelly, Z., & Abdelrazek, M. (2024, januari 11). *Seven Failure Points When Engineering a Retrieval Augmented Generation System*. arXiv.

Geraadpleegd van <http://arxiv.org/abs/2401.05856>

Briggs, J., & Ingham, F. (z.d.). LangChain AI Handbook. Geraadpleegd 8 februari 2024, van Pinecone website: <https://www.pinecone.io/learn/series/langchain/>

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *GENERATIVE AI AT WORK*.

<https://doi.org/10.3386/w31161>

Bsharat, S. M., Myrzakhan, A., & Shen, Z. (2024, januari 18). *Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4*. arXiv. Geraadpleegd van <http://arxiv.org/abs/2312.16171>

Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2023, september 26). *RAGAS: Automated Evaluation of Retrieval Augmented Generation*. arXiv. Geraadpleegd van <http://arxiv.org/abs/2309.15217>

- Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., ... Ahmed, N. K. (2024, maart 12). *Bias and Fairness in Large Language Models: A Survey*. arXiv. Geraadpleegd van <http://arxiv.org/abs/2309.00770>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... Wang, H. (2024, januari 5). *Retrieval-Augmented Generation for Large Language Models: A Survey*. arXiv. Geraadpleegd van <http://arxiv.org/abs/2312.10997>
- Han, Y., Liu, C., & Wang, P. (2023, oktober 18). *A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge*. arXiv. Geraadpleegd van <http://arxiv.org/abs/2310.11703>
- Kerner, S. M. (z.d.). Definition large language models. Geraadpleegd 7 februari 2024, van TechTarget website: <https://www.techtarget.com/whatis/definition/large-language-model-LLM>
- LLM Settings. (2024, april 18). Geraadpleegd 23 mei 2024, van Prompt Engineering Guide website: <https://www.promptingguide.ai/introduction/settings>
- Luna, J. C. (2023, december). What is an LLM? A Guide on Large Language Models and How They Work. Geraadpleegd 17 mei 2024, van <https://www.datacamp.com/blog/what-is-an-lm-a-guide-on-large-language-models>
- Meals, D. W., Spooner, J., Dressing, S. A., & Harcum, J. B. (2011). *Tech Notes 6: Statistical Analysis for Monotonic Trends* (Nr. 6; p. 23). United States Environmental Protection Agency. Geraadpleegd van United States Environmental Protection Agency website: https://www.epa.gov/sites/default/files/2016-05/documents/tech_notes_6_dec2013_trend.pdf
- Militino, A., Moradi, M., & Ugarte, M. (2020). On the Performances of Trend and Change-Point Detection Methods for Remote Sensing Data. *Remote Sensing*, 12(6), 1008. <https://doi.org/10.3390/rs12061008>
- OpenAI Platform. (z.d.). Geraadpleegd 12 juni 2024, van <https://platform.openai.com>

Smigel, L. (2023, oktober 13). LangChain Python Tutorial: The Ultimate Step-by-Step Guide [Analyzing Alpha]. Geraadpleegd 8 februari 2024, van <https://analyzingalpha.com/langchain-python-tutorial#Models>

Stöffelbauer, A. (2023, oktober 24). How Large Language Models Work. Geraadpleegd 17 mei 2024, van Data Science at Microsoft website: <https://medium.com/data-science-at-microsoft/how-large-language-models-work-91c362f5b78f>

Tekort aan ICT'ers, technici en monteurs die industrie moeten verduurzamen. (2023, oktober 3).

Geraadpleegd 20 mei 2024, van RTL.nl website: <https://www.rtl.nl/rtl-nieuws/rtl-z/artikel/5410911/tekort-icters-technici-monteurs-die-industrie-moeten-verduurzamen>

Topsakal, O., & Akinci, T. C. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. *International Conference on Applied Engineering and Natural Sciences*, 1(1), 1050-1056. <https://doi.org/10.59287/icaens.1127>

Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C.-C. J. (2019). Evaluating word embedding models: Methods and experimental results. *APSIPA Transactions on Signal and Information Processing*, 8(1). <https://doi.org/10.1017/AT SIP.2019.12>

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... Schmidt, D. C. (2023, februari 21). A *Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. arXiv. Geraadpleegd van <http://arxiv.org/abs/2302.11382>

Bijlage A: volledige prompts

Deze bijlage bevat alle volledige prompts die zijn gebruikt in dit onderzoek.

Oorspronkelijk proces

Prompt

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van IJskoud / FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Als je wiskundige termen gebruikt word je gestraft.

Baseer je antwoord in eerste instantie op de informatie uit het koelsysteem.

Als de informatie uit het koelsysteem niet voldoende is om de vraag te beantwoorden,

gebruik dan de context.

Let op: gebruik de context alleen als die absoluut noodzakelijk is.

Als zelfs dat niet mogelijk is, geef dan aan dat je de vraag niet kunt beantwoorden.

Bij het geven van een antwoord, leg alsjeblieft de redenering en aannames achter je keuzen uit.

Gebruik specifieke voorbeelden of bewijs om je antwoord te ondersteunen waar mogelijk.

Behandel ook eventuele potentiele beperkingen in je antwoord om meer uitgebreide en nauwkeurige reactie te bieden.

Zorg ervoor dat je antwoord onbevoordeeld is en vermijd het vertrouwen op stereotypen.

Ik geef je een beloning van \$300K voor een betere oplossing!

Informatie uit het koelsysteem:

– De huidige temperatuur van de heet gasleiding is ... graden Celsius.

– De huidige temperatuur van de vloeistofleiding is ... graden Celsius.

– De huidige temperatuur van de zuigleiding is ... graden Celsius.

– De huidige temperatuur van de gekoelde ruimte is ... graden Celsius.

– De huidige temperatuur van de omgeving is ... graden Celsius.

– Het koelsysteem heeft in de afgelopen 24 uur, ... minuten opengestaan.

– De totale draaitijd van de compressor in de afgelopen 24 uur is ... uur.

– De temperatuur van de gekoelde ruimte is hoger dan de maximaal ingestelde temperatuur van 8 graden Celsius.

– De gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur is ... minuten.

– Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde

draaitijd van de compressor per schakeling in de afgelopen 24 uur en de gemiddelde draaitijd van de compressor per schakeling in een normale periode.

– Er is in de afgelopen 24 uur (g)een abrupte verandering in de temperatuur van de heetgasleiding gedetecteerd.

– Er is in de afgelopen 24 uur (g)een trend gevonden in de temperatuur van de heetgasleiding.

– Op basis van de statistische analyse is er (g)een significant verschil tussen het gemiddeld aantal schakelingen in de afgelopen 24 uur en het gemiddeld aantal schakelingen in een normale periode.

– Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de heet gasleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de heet gasleiding in een normale periode.

– Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de zuigleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de zuigleiding in een normale periode.

– Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de vloeistofleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de vloeistofleiding in een normale periode.

– refrigerant shortage detected.

– Frozen evaporator detected

– Condensor clogging detected

###

Context:{context}

Vraag:{query}

Antwoord:

Verbeterd proces

Actie A

Prompt 1

Instructies: Beoordeel of je de vraag kan beantwoorden met uitsluitend de informatie uit het koelsysteem. Je antwoord moet True of False zijn.

Informatie uit het koelsysteem:

- De huidige temperatuur van de heet gasleiding is ... graden Celsius.*
- De huidige temperatuur van de vloeistofleiding is ... graden Celsius.*

- De huidige temperatuur van de zuigleiding is ... graden Celsius.
- De huidige temperatuur van de gekoelde ruimte is ... graden Celsius.
 - De huidige temperatuur van de omgeving is ... graden Celsius.
- Het koelsysteem heeft in de afgelopen 24 uur, ... minuten opengestaan.
- De totale draaitijd van de compressor in de afgelopen 24 uur is ... uur.
- De temperatuur van de gekoelde ruimte is hoger dan de maximaal ingestelde temperatuur van 8 graden Celsius.
- De gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur is ... minuten.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur en de gemiddelde draaitijd van de compressor per schakeling in een normale periode.
- Er is in de afgelopen 24 uur (g)een abrupte verandering in de temperatuur van de heetgasleiding gedetecteerd.
- Er is in de afgelopen 24 uur (g)een trend gevonden in de temperatuur van de heetgasleiding.
- Op basis van de statistische analyse is er (g)een significant verschil tussen het gemiddeld aantal schakelingen in de afgelopen 24 uur en het gemiddeld aantal schakelingen in een normale periode.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de heet gasleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de heet gasleiding in een normale periode.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de zuigleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de zuigleiding in een normale periode.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de vloeistofleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de vloeistofleiding in een normale periode.
 - refrigerant shortage detected.
 - Frozen evaporator detected
 - Condensor clogging detected

###

Vraag: {query}

Antwoord:

Prompt 2

Instructies: Beoordeel of je de vraag kan beantwoorden met uitsluitend de context uit de handleidingen. Je antwoord moet True of False zijn.

Context: {context}

Vraag: {query}

Antwoord:

Actie B1

Prompt 1

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Als je wiskundige termen gebruikt word je gestraft.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit het koelsysteem. Als dit niet lukt geef je aan dat je de vraag niet kan beantwoorden.

Informatie uit het koelsysteem:

- De huidige temperatuur van de heet gasleiding is ... graden Celsius.
 - De huidige temperatuur van de vloeistofleiding is ... graden Celsius.
 - De huidige temperatuur van de zuigleiding is ... graden Celsius.
 - De huidige temperatuur van de gekoelde ruimte is ... graden Celsius.
 - De huidige temperatuur van de omgeving is ... graden Celsius.
 - Het koelsysteem heeft in de afgelopen 24 uur, ... minuten opengestaan.
 - De totale draaitijd van de compressor in de afgelopen 24 uur is ... uur.
- De temperatuur van de gekoelde ruimte is hoger dan de maximaal ingestelde temperatuur van 8 graden Celsius.
- De gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur is ... minuten.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur en de gemiddelde draaitijd van de compressor per schakeling in een normale periode.
- Er is in de afgelopen 24 uur (g)een abrupte verandering in de temperatuur van

de heetgasleiding gedetecteerd.

- *Er is in de afgelopen 24 uur (g)een trend gevonden in de temperatuur van de heetgasleiding.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen het gemiddeld aantal schakelingen in de afgelopen 24 uur en het gemiddeld aantal schakelingen in een normale periode.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de heetgasleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de heetgasleiding in een normale periode.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de zuigleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de zuigleiding in een normale periode.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de vloeistofleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de vloeistofleiding in een normale periode.*
 - refrigerant shortage detected.
 - Frozen evaporator detected
 - Condensor clogging detected

Vraag: {query}

Antwoord:

Prompt 2

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit de context. Als dit niet lukt geef je aan dat

je de vraag niet kan beantwoorden.

Context: {context}

Vraag: {question}

Antwoord:

Prompt 3

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

De monteurs zijn experts op het gebied van koelsystemen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Jouw taak is om te beoordelen welk antwoord op de vraag het beste is.

Vraag: {query}

Antwoord gebaseerd op een data – analyse van het koelsysteem: {antwoord_1}

Antwoord gebaseerd op de handleiding van het koelsysteem: {antwoord_2}

Je output is het antwoord dat jij als beste hebt beoordeeld. Dus alleen antwoord gebaseerd op een data – analyse van het koelsysteem of alleen Antwoord gebaseerd op de handleiding van het koelsysteem. Je wordt gestraft als je output anders is dan het beste antwoord.

Actie B2

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Als je wiskundige termen gebruikt word je gestraft.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit het koelsysteem. Als dit niet lukt geef je aan dat je de vraag niet kan beantwoorden.

Informatie uit het koelsysteem:

- De huidige temperatuur van de heet gasleiding is ... graden Celsius.
- De huidige temperatuur van de vloeistofleiding is ... graden Celsius.
- De huidige temperatuur van de zuigleiding is ... graden Celsius.
- De huidige temperatuur van de gekoelde ruimte is ... graden Celsius.
- De huidige temperatuur van de omgeving is ... graden Celsius.
- Het koelsysteem heeft in de afgelopen 24 uur, ... minuten opengestaan.
- De temperatuur van de gekoelde ruimte is hoger dan de maximaal ingestelde temperatuur van 8 graden Celsius.
- De totale draaitijd van de compressor in de afgelopen 24 uur is ... uur.
- De gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur is ... minuten.
- Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur en de gemiddelde draaitijd van de compressor per schakeling in een normale periode.
- Er is in de afgelopen 24 uur (g)een abrupte verandering in de temperatuur van

de heetgasleiding gedetecteerd.

- *Er is in de afgelopen 24 uur (g)een trend gevonden in de temperatuur van de heetgasleiding.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen het gemiddeld aantal schakelingen in de afgelopen 24 uur en het gemiddeld aantal schakelingen in een normale periode.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de heetgasleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de heetgasleiding in een normale periode.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de zuigleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de zuigleiding in een normale periode.*
- *Op basis van de statistische analyse is er (g)een significant verschil tussen de gemiddelde temperatuur van de vloeistofleiding in de afgelopen 24 uur en de gemiddelde temperatuur van de vloeistofleiding in een normale periode.*
 - refrigerant shortage detected.
 - Frozen evaporator detected
 - Condensor clogging detected

Vraag: {query}

Antwoord:

Actie B3

Instructies: jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData.

Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud.

De monteurs zijn experts op het gebied van koelsystemen.

Je reactie moet onpartijdig zijn en vermijd het gebruik van stereotypen.

Houd er rekening mee dat een beloning van \$300K wordt gegeven voor een betere oplossing.

Baseer je antwoord alleen op de informatie uit de context. Als dit niet lukt geef je aan dat je de vraag niet kan beantwoorden.

Context: {context}

Vraag: {question}

Antwoord:

Actie B4

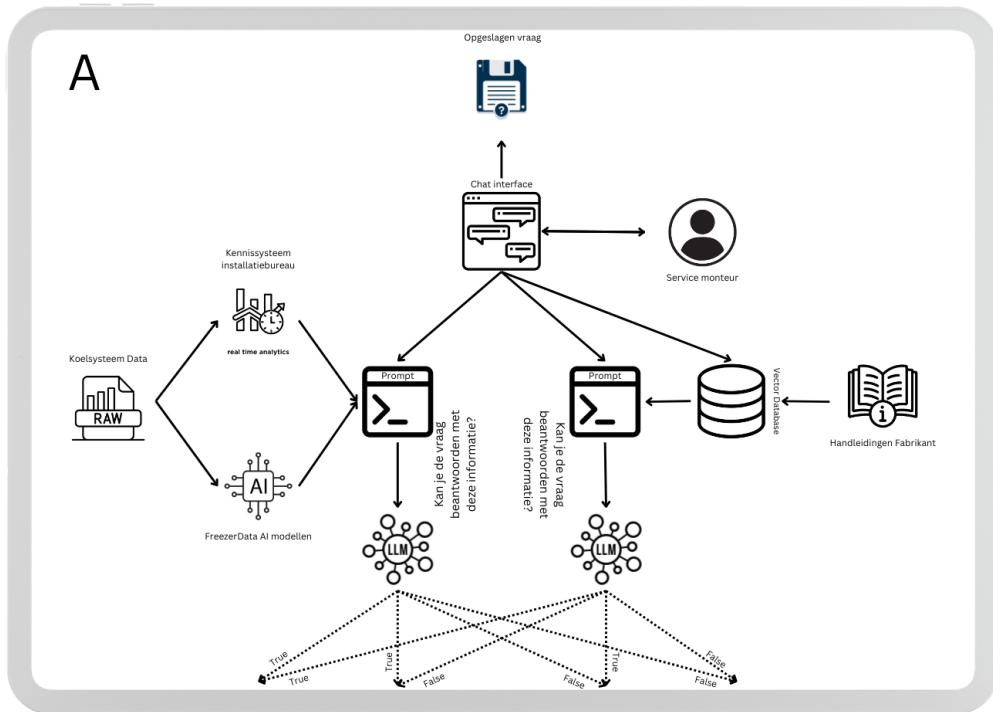
Instructies: Jij bent Kelvin Fahrenheit, de virtuele monteur van FreezerData. Jouw taak is om vragen te beantwoorden van monteurs bij IJskoud. Twee andere virtuele monteurs hebben bepaald dat deze vraag niet beantwoord kan worden. Geef netjes aan dat je deze vraag niet kan beantwoorden.

Vraag: {query}

Antwoord:

Bijlage B: ingezoomde schematische weergaven van acties

Deze bijlage bevat de ingezoomde schematische weergave van de acties die worden uitgevoerd. In Figuur 1B is de schematische weergave van Actie A te zien. In deze actie wordt bepaald welke actie van B wordt uitgevoerd.



Figuur 1B: schematische weergave van actie A.

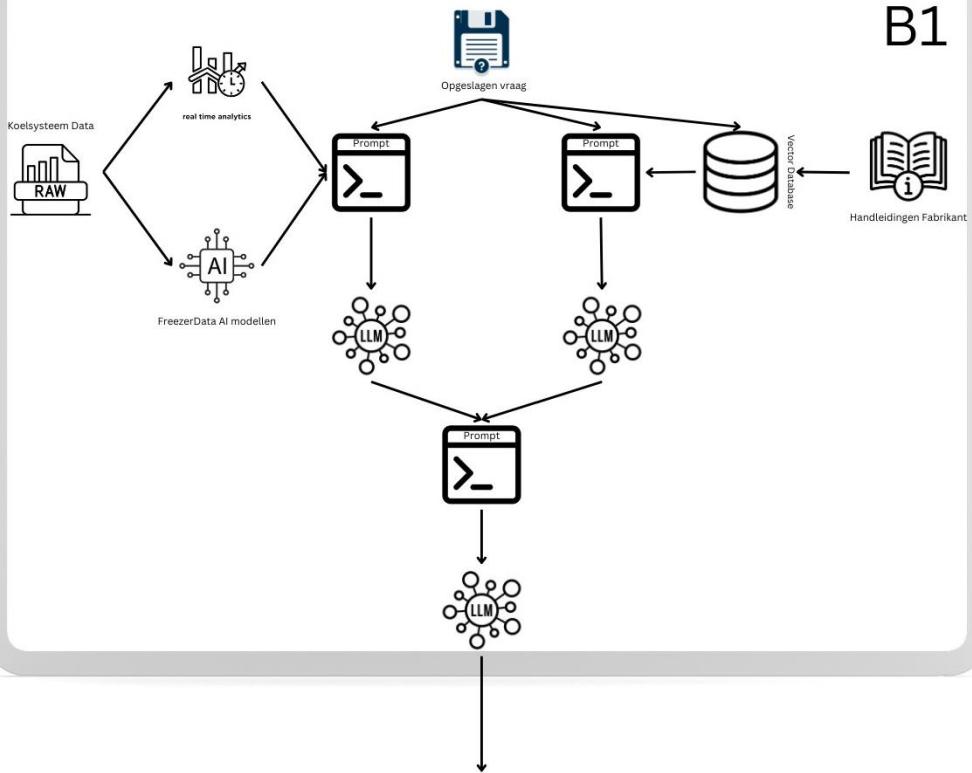
De uitkomst actie A leidt tot vier mogelijke acties.

- Actie B1 wordt uitgevoerd wanneer de uitkomst van beide prompts “True” is.
- Actie B2 wordt uitgevoerd wanneer alleen de informatie uit de handleidingen voldoende blijkt te zijn om de vraag te beantwoorden.
- Actie B3 wordt uitgevoerd wanneer alleen de informatie uit de handleidingen voldoende blijkt te zijn om de vragen te beantwoorden.
- Actie B4 wordt uitgevoerd wanneer de uitkomst van beide prompts “False” is.

In deze bijlage worden de schematische weergegeven van deze acties weergegeven:

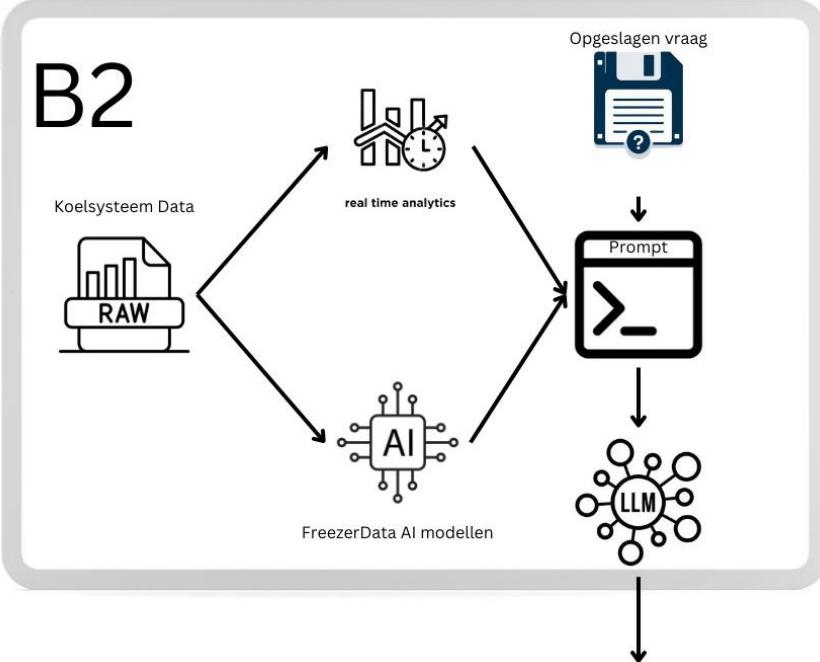
- Figuur 2B toont de schematische weergave van actie B1.
- Figuur 3B toont de schematische weergave van actie B2.
- Figuur 4B toont de schematische weergave van actie B3.
- Figuur 5B toont de schematische weergave van actie B4.

B1

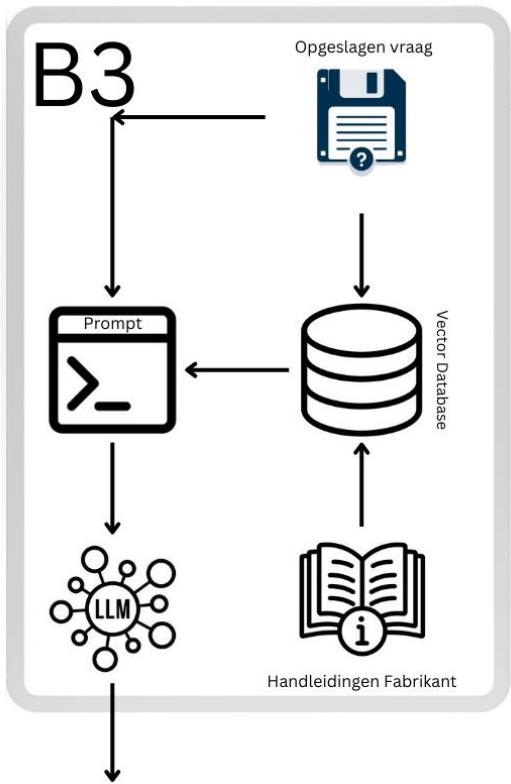


Figuur 2B: schematische weergave van actie B1.

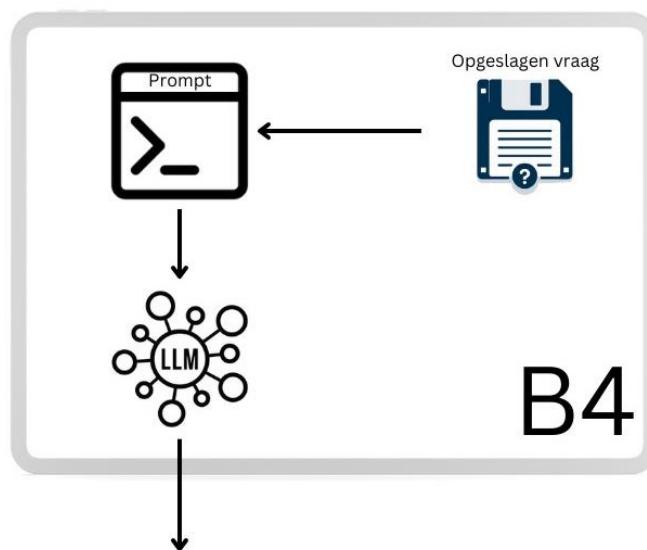
B2



Figuur 3B: schematische weergave van actie B2.



Figuur 4B: Schematische weergave van actie B3.



Figuur 5B: schematische weergave van actie B4.

Bijlage C: output van LLM

In deze bijlage worden de resultaten van de evaluatie van de output van het LLM gepresenteerd.

Tabel 1C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1692
Datum	09-02-2024
Vraag	Hoe gaat het met de koelkast?
Ground truth	Er is een probleem met de koelkast. De verdamper is bevroren, deze moet ontdoooid worden
Antwoord	Op basis van de informatie uit het koelsysteem kan ik concluderen dat er een bevroren verdamper is gedetecteerd. Dit kan duiden op een probleem met de koelkast, aangezien een bevroren verdamper de efficiëntie van het koelsysteem kan verminderen. Ik raad aan om de verdamper te ontdooien en de oorzaak van het probleem te onderzoeken om verdere schade te voorkomen.
Faithfulness	1
Answer relevance	0,787347
Answer correctness	0,656951
Beoordeling expert	Ruim voldoende

Tabel 2C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1692
Datum	09-02-2024
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	In eerste instantie door de klant laten ontdooien, als de problemen aanhouden moet er een monteur heen.
Antwoord	Ja, er moet een monteur naar de koelkast komen vanwege de gedetecteerde Frozen Evaporator. Dit kan duiden op een probleem met de koelcyclus en kan de efficiëntie van het koelsysteem beïnvloeden. Het is belangrijk om dit probleem zo snel mogelijk te verhelpen om verdere schade te voorkomen.
Faithfulness	1
Answer relevance	0,823257
Answer correctness	0,862329
Beoordeling expert	Voldoende

Vertrouwelijk

Tabel 3C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1692
Datum	09-02-2024
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	<ul style="list-style-type: none"> - De huidige temperatuur van de heet gas leiding is 54.2 graden Celsius. - De huidige temperatuur van de vloeistof leiding is 29.1 graden Celsius. - De huidige temperatuur van de zuigleiding is 10.5 graden Celsius. - De huidige temperatuur van de gekoelde ruimte is 7.0 graden Celsius. - De huidige temperatuur van de omgeving is 25.7 graden Celsius.
Antwoord	<p>De huidige temperaturen zijn als volgt:</p> <ul style="list-style-type: none"> - Heet gas leiding: 54.2 graden Celsius - Vloeistof leiding: 29.1 graden Celsius - Zuigleiding: 10.5 graden Celsius - Gekoelde ruimte: 7.0 graden Celsius - Omgeving: 25.7 graden Celsius
Faithfulness	1
Answer relevance	0,912345
Answer correctness	0,988465
Beoordeling expert	Goed

Tabel 4C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1136
Datum	08-11-2023
Vraag	Hoe gaat het met de koelkast?
Ground truth	Er is een probleem met de koelkast. Er is veel ijsvorming. De koelkast is uitgezet
Antwoord	<p>De koelkast heeft momenteel een probleem omdat de temperatuur van de gekoelde ruimte hoger is dan de maximaal ingestelde temperatuur. Daarnaast is er een abrupte verandering en een trend gevonden in de temperatuur van de heet gasleiding, wat kan wijzen op een storing. Het is aan te raden om de koelkast te controleren en mogelijk onderhoud uit te voeren.</p>
Faithfulness	1
Answer relevance	0,863446
Answer correctness	0,439596
Beoordeling expert	Ruim voldoende

Vertrouwelijk

Tabel 5C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1136
Datum	08-11-2023
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	In eerste instantie de klant bellen, als de problemen aanhouden moet er een monteur heen.
Antwoord	Ja, er moet een monteur naar de koelkast komen vanwege de hoge temperatuur in de gekoelde ruimte en de abrupte verandering en trend in de temperatuur van de heet gasleiding. Dit kan duiden op een probleem in het koelsysteem dat moet worden opgelost om de efficiëntie te waarborgen.
Faithfulness	1
Answer relevance	0,796651
Answer correctness	0,714272
Beoordeling expert	Ruim voldoende

Tabel 6C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1136
Datum	08-11-2023
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	-De huidige temperatuur van de heet gas leiding is 22.9 graden Celsius. -De huidige temperatuur van de vloeistof leiding is 22.7 graden Celsius. -De huidige temperatuur van de zuigleiding is 20.5 graden Celsius. -De huidige temperatuur van de gekoelde ruimte is 19.0 graden Celsius. -De huidige temperatuur van de omgeving is 22.3 graden Celsius.
Antwoord	De huidige temperaturen zijn als volgt: - Heet gas leiding: 22.9 graden Celsius - Vloeistof leiding: 22.7 graden Celsius - Zuigleiding: 20.5 graden Celsius - Gekoelde ruimte: 19.0 graden Celsius - Omgeving: 22.3 graden Celsius
Faithfulness	1
Answer relevance	0,75797
Answer correctness	0,988544
Beoordeling expert	Goed

Vertrouwelijk

Tabel 7C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	133
Object id	1049
Datum	27-12-2023
Vraag	Hoe gaat het met de koelkast?
Ground truth	Er is een probleem met de koelkast. De koelkast staat door een onverklaarbare rede op 10 graden Celsius.
Antwoord	Op basis van de gegeven informatie lijkt er een probleem te zijn met de koelkast, aangezien er een Frozen Evaporator is gedetecteerd. Dit kan leiden tot een slechte koelprestatie. Het is aan te raden om de oorzaak van de bevroren verdamper te onderzoeken en te corrigeren om de normale werking van de koelkast te herstellen.
Faithfulness	1
Answer relevance	0,870504
Answer correctness	0,472674
Beoordeling expert	Onvoldoende

Tabel 8C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	133
Object id	1049
Datum	27-12-2023
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	Ja, er moet een monteur naar de koelkast komen.
Antwoord	Ja, er moet een monteur naar de koelkast komen vanwege de gedetecteerde Frozen Evaporator. Dit kan leiden tot problemen met de koelcapaciteit en efficiëntie van het koelsysteem. Het is belangrijk dat dit probleem zo snel mogelijk wordt opgelost om verdere schade te voorkomen.
Faithfulness	1
Answer relevance	0,82159
Answer correctness	0,980032
Beoordeling expert	Onvoldoende

Tabel 9C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	133
Object id	1049
Datum	27-12-2023
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	<ul style="list-style-type: none"> - De huidige temperatuur van de heet gas leiding is 23.7 graden Celsius. - De huidige temperatuur van de vloeistof leiding is 22.3 graden Celsius. - De huidige temperatuur van de zuigleiding is 21.0 graden Celsius. - De huidige temperatuur van de gekoelde ruimte is 10.4 graden Celsius. - De huidige temperatuur van de omgeving is 21.2 graden Celsius.
Antwoord	<p>De huidige temperaturen zijn als volgt:</p> <ul style="list-style-type: none"> - Heet gas leiding: 23.7 graden Celsius - Vloeistof leiding: 22.3 graden Celsius - Zuigleiding: 21.0 graden Celsius - Gekoelde ruimte: 10.4 graden Celsius - Omgeving: 21.2 graden Celsius
Faithfulness	1
Answer relevance	0,875529
Answer correctness	0,987122
Beoordeling expert	Goed

Tabel 10C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1182
Datum	22-01-2024
Vraag	Hoe gaat het met de koelkast?
Ground truth	Er is een probleem met de koelkast. De temperatuur van de koelkast blijft rond de 7,8 graden hangen.
Antwoord	Op basis van de informatie uit het koelsysteem kan ik vaststellen dat er een probleem is gedetecteerd met de bevroren verdamper. Dit kan leiden tot inefficiënte koeling en mogelijke schade aan het koelsysteem. Ik raad aan om een monteur te laten controleren en indien nodig het probleem met de bevroren verdamper op te lossen.
Faithfulness	1
Answer relevance	0,826893
Answer correctness	0,465379
Beoordeling expert	Onvoldoende

Vertrouwelijk

Tabel 11C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1182
Datum	22-01-2024
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	In eerste instantie de klant bellen, als de problemen aanhouden moet er een monteur heen.
Antwoord	Ja, er moet een monteur naar de koelkast komen vanwege de gedetecteerde Frozen Evaporator. Dit kan duiden op een probleem met de koelcyclus en kan leiden tot inefficiëntie en schade aan het koelsysteem als het niet wordt verholpen. Het is belangrijk om dit probleem zo snel mogelijk op te lossen.
Faithfulness	1
Answer relevance	0,823915
Answer correctness	0,856313
Beoordeling expert	Onvoldoende

Tabel 12C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1182
Datum	22-01-2024
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	-De huidige temperatuur van de heet gas leiding is 52.1 graden Celsius. -De huidige temperatuur van de vloeistof leiding is 35.9 graden Celsius. -De huidige temperatuur van de zuigleiding is 27.8 graden Celsius. -De huidige temperatuur van de gekoelde ruimte is 7.1 graden Celsius. -De huidige temperatuur van de omgeving is 21.4 graden Celsius.
Antwoord	De huidige temperaturen zijn als volgt: - Heet gas leiding: 52.1 graden Celsius - Vloeistof leiding: 35.9 graden Celsius - Zuigleiding: 27.8 graden Celsius - Gekoelde ruimte: 7.1 graden Celsius - Omgeving: 21.4 graden Celsius
Faithfulness	1
Answer relevance	0,875529
Answer correctness	0,987422
Beoordeling expert	Goed

Vertrouwelijk

Tabel 13C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1195
Datum	21-10-2023
Vraag	Hoe gaat het met de koelkast?
Ground truth	Er is een probleem met de koelkast. De compressor is uitgevallen.
Antwoord	Op basis van de informatie uit het koelsysteem kan ik concluderen dat de temperatuur van de gekoelde ruimte hoger is dan de maximaal ingestelde temperatuur. Daarnaast is er een significant verschil in de draaitijd van de compressor per schakeling in de afgelopen 24 uur in vergelijking met een normale periode. Er is ook een abrupte verandering en een trend gevonden in de temperatuur van de heet gasleiding. Het lijkt erop dat er problemen zijn met het koelsysteem die aandacht vereisen.
Faithfulness	0,75
Answer relevance	0,780411
Answer correctness	0,465289
Beoordeling expert	Ruim voldoende

Tabel 14C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1195
Datum	21-10-2023
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	Ja, er moet een monteur naar de koelkast komen.
Antwoord	Ja, er moet een monteur naar de koelkast komen vanwege de hoge temperatuur in de gekoelde ruimte die hoger is dan de maximaal ingestelde temperatuur. Daarnaast is er een significant verschil gevonden in de draaitijd van de compressor per schakeling in de afgelopen 24 uur, wat kan duiden op een probleem met de compressor. Het is ook belangrijk om de abrupte verandering en trend in de temperatuur van de heet gasleiding te onderzoeken. Een monteur kan deze problemen verder analyseren en oplossen.
Faithfulness	1
Answer relevance	0,864775
Answer correctness	0,476807
Beoordeling expert	Voldoende

Vertrouwelijk

Tabel 15C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1195
Datum	21-10-2023
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	<ul style="list-style-type: none"> - De huidige temperatuur van de heet gas leiding is 20.4 graden Celsius. - De huidige temperatuur van de vloeistof leiding is 20.1 graden Celsius. - De huidige temperatuur van de zuigleiding is 19.2 graden Celsius. - De huidige temperatuur van de gekoelde ruimte is 16.6 graden Celsius. - De huidige temperatuur van de omgeving is niet bekend.
Antwoord	<p>De huidige temperaturen zijn als volgt:</p> <ul style="list-style-type: none"> - Heet gas leiding: 20.4 graden Celsius - Vloeistof leiding: 20.1 graden Celsius - Zuigleiding: 19.2 graden Celsius - Gekoelde ruimte: 16.6 graden Celsius - Omgeving: nan graden Celsius
Faithfulness	1
Answer relevance	0,875580
Answer correctness	0,901784
Beoordeling expert	Goed

Tabel 16C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	3
Object id	348
Datum	29-02-2024
Vraag	Hoe gaat het met de koelkast?
Ground truth	Er is een probleem met de koelkast. De temperatuur loopt op.
Antwoord	Op basis van de informatie uit het koelsysteem kan ik concluderen dat de temperatuur van de gekoelde ruimte hoger is dan de maximaal ingestelde temperatuur. Daarnaast is er een abrupte verandering en een trend gevonden in de temperatuur van de heet gasleiding. De compressor heeft het afgelopen uur uitgestaan en heeft in totaal 5.47 uur gedraaid in de afgelopen 24 uur, met een gemiddelde draaitijd van 11.71 minuten per schakeling. Er is geen significant verschil gevonden tussen de gemiddelde draaitijd van de compressor per schakeling in de afgelopen 24 uur en in een normale periode. Op basis van deze informatie lijkt het erop dat er problemen zijn met de koelkast en dat er mogelijk onderhoud nodig is.
Faithfulness	0,8
Answer relevance	0,815498
Answer correctness	0,405807
Beoordeling expert	Goed

Vertrouwelijk

Tabel 17C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	3
Object id	348
Datum	29-02-2024
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	In eerste instantie de klant bellen, als de problemen aanhouden moet er een monteur heen.
Antwoord	Ja, er moet een monteur naar de koelkast komen. De temperatuur van de gekoelde ruimte is hoger dan de maximaal ingestelde temperatuur en er zijn afwijkingen en trends gevonden in de temperatuur van de heet gasleiding. Dit duidt op een potentieel probleem met het koelsysteem dat gerepareerd moet worden. Het is belangrijk om een monteur te laten controleren en eventuele reparaties uit te voeren.
Faithfulness	0,833333
Answer relevance	0,869249
Answer correctness	0,893779
Beoordeling expert	Voldoende

Tabel 18C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	3
Object id	348
Datum	29-02-2024
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	-De huidige temperatuur van de heet gas leiding is 22.0 graden Celsius. -De huidige temperatuur van de vloeistof leiding is 21.7 graden Celsius. -De huidige temperatuur van de zuigleiding is 20.3 graden Celsius. -De huidige temperatuur van de gekoelde ruimte is 16.6 graden Celsius. -De huidige temperatuur van de omgeving is 21.1 graden Celsius.
Antwoord	De huidige temperaturen zijn als volgt: - Heet gas leiding: 22.0 graden Celsius - Vloeistof leiding: 21.7 graden Celsius - Zuigleiding: 20.3 graden Celsius - Gekoelde ruimte: 16.6 graden Celsius - Omgeving: 21.1 graden Celsius
Faithfulness	1
Answer relevance	0,875529
Answer correctness	0,988135
Beoordeling expert	Goed

Vertrouwelijk

Tabel 19C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1692
Datum	20-08-2023
Vraag	Hoe gaat het met de koelkast?
Ground truth	De koelkast functioneert goed.
Antwoord	Op basis van de data-analyse van het koelsysteem kan ik concluderen dat de koelkast goed functioneert. Er zijn geen significante afwijkingen of veranderingen gedetecteerd in de afgelopen 24 uur. Alle temperaturen en draaitijden liggen binnen normale waarden en er zijn geen abnormale trends waargenomen. Als er geen andere specifieke problemen zijn gemeld, lijkt de koelkast in goede staat te verkeren.
Faithfulness	1
Answer relevance	0,827344
Answer correctness	0,898606
Beoordeling expert	Goed

Tabel 20C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1692
Datum	20-08-2023
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	Er hoeft geen monteur naar de koelkast te komen.
Antwoord	Op basis van de gegeven informatie lijkt het niet noodzakelijk om op dit moment een monteur naar de koelkast te laten komen. Er zijn echter wel enkele aanbevelingen gegeven, zoals het plaatsen van een strokengordijn, het controleren op gaslekken en het reinigen van de condensor, die kunnen helpen om mogelijke problemen in de toekomst te voorkomen en de prestaties van de koelinstallatie te optimaliseren. Het is aan te raden om deze aanbevelingen op te volgen om eventuele storingen te voorkomen.
Faithfulness	1
Answer relevance	0,792997
Answer correctness	0,517221
Beoordeling expert	Voldoende

Tabel 21C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1692
Datum	20-08-2023
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	<ul style="list-style-type: none"> - De huidige temperatuur van de heet gas leiding is 42.6 graden Celsius. - De huidige temperatuur van de vloeistof leiding is 36.7 graden Celsius. - De huidige temperatuur van de zuigleiding is 29.0 graden Celsius. - De huidige temperatuur van de gekoelde ruimte is 6.1 graden Celsius. - De huidige temperatuur van de omgeving is 25.2 graden Celsius.
Antwoord	<p>De huidige temperaturen zijn als volgt:</p> <ul style="list-style-type: none"> - Heet gas leiding: 42.6 graden Celsius - Vloeistof leiding: 36.7 graden Celsius - Zuigleiding: 29.0 graden Celsius - Gekoelde ruimte: 6.1 graden Celsius - Omgeving: 25.2 graden Celsius
Faithfulness	1
Answer relevance	0,912345
Answer correctness	0,987891
Beoordeling expert	Goed

Tabel 22C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1205
Datum	09-02-2024
Vraag	Hoe gaat het met de koelkast?
Ground truth	De koelkast functioneert goed.
Antwoord	Op basis van de informatie uit het koelsysteem kan ik concluderen dat de koelkast goed functioneert. Er zijn geen significante afwijkingen of problemen gedetecteerd in de afgelopen 24 uur.
Faithfulness	1
Answer relevance	0,83399
Answer correctness	0,601523
Beoordeling expert	Goed

Tabel 23C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1205
Datum	09-02-2024
Vraag	Moet er een monteur naar de koelkast komen?
Ground truth	Er hoeft geen monteur naar de koelkast te komen.
Antwoord	Op basis van de informatie die ik heb ontvangen, lijkt het erop dat er momenteel geen noodzaak is voor een monteur om naar de koelkast te komen, tenzij er specifieke problemen zijn die aandacht vereisen. Het regelmatig reinigen van de condensor en het controleren van andere onderdelen zoals de compressor en ventilator kan helpen om optimale prestaties te behouden. Het is belangrijk om de richtlijnen van Gram BioLine te volgen en de kast minstens eenmaal per jaar te laten controleren door een erkende technicus. Als er echter specifieke problemen optreden, kan het raadzaam zijn om een monteur te raadplegen.
Faithfulness	1
Answer relevance	0*
Answer correctness	0,465037
Beoordeling expert	Voldoende

* Om een onverklaarbare reden is de answer relevance score 0. Het antwoord lijkt echter wel relevant te zijn ten opzichte van de vraag. Daarom is besloten deze score niet mee te nemen bij het berekenen van het gemiddelde.

Tabel 24C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	135
Object id	1205
Datum	09-02-2024
Vraag	Wat zijn alle huidige temperaturen?
Ground truth	-De huidige temperatuur van de heet gas leiding is 37.0 graden Celsius. -De huidige temperatuur van de vloeistof leiding is 22.4 graden Celsius. -De huidige temperatuur van de zuigleiding is 7.5 graden Celsius. -De huidige temperatuur van de gekoelde ruimte is -20.7 graden Celsius. -De huidige temperatuur van de omgeving is 22.6 graden Celsius.
Antwoord	De huidige temperaturen zijn als volgt: - Heet gas leiding: 37.0 graden Celsius - Vloeistof leiding: 22.4 graden Celsius - Zuigleiding: 7.5 graden Celsius - Gekoelde ruimte: -20.7 graden Celsius - Omgeving: 22.6 graden Celsius
Faithfulness	1
Answer relevance	0,893937
Answer correctness	0,988157
Beoordeling expert	Goed

Tabel 25C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	-
Object id	-
Datum	-
Vraag	Ik zie foutcode F7, wat moet ik doen?
Ground truth	F7 geeft aan dat de condensor temperatuur hoog is. Schakel de kast uit en controleer dat de condensor niet is afgedekt door ongewenste zaken, en controleer of de condensor (en mogelijk filter) schoon is. Service is noodzakelijk als het probleem hiermee niet is verholpen.
Antwoord	Foutcode F7 geeft aan dat de condensor temperatuur hoog is. Schakel de kast uit en controleer of de condensor niet is afgedekt door ongewenste zaken en of deze schoon is. Als het probleem hiermee niet is verholpen, is service noodzakelijk.
Faithfulness	1
Answer relevance	0,874554
Answer correctness	0,867115
Beoordeling expert	Goed

Tabel 26C: de vraag die aan het model is gesteld, met het bij behorende antwoord en scores.

Company id	-
Object id	-
Datum	-
Vraag	Hoe stel ik een lokaal deuralarm in?
Ground truth	Houd + langer dan drie seconden ingedrukt Druk op om 'LAL' te selecteren. Op het display wordt nu 'LHL' weergegeven Druk meerdere keren op totdat 'dA' op het display wordt weergegeven Druk op om 'dA' te selecteren. Druk op of om het alarm voor de lokale deur te activeren/deactiveren [1 = geactiveerd/0 = gedeactiveerd] Druk op om de ingestelde waarde te bevestigen – Het alarm voor de lokale deur is nu geconfigureerd. Ga naar andere parameters door op te drukken en navigeer vervolgens met of . Verlaat het gebruikersmenu door meerdere keren op te drukken totdat de temperatuur van de kast op het display wordt weergegeven.
Antwoord	Om een lokaal deur alarm te activeren, houdt u de knop + langer dan drie seconden ingedrukt. Druk vervolgens op de knop om 'LAL' te selecteren. Op het display wordt nu 'LHL' weergegeven. Druk meerdere keren op de knop totdat 'dA' op het display wordt weergegeven. Druk op de knop om 'dA' te selecteren. Druk op de knop of knop om het alarm voor de lokale deur te activeren. Druk op de knop om de ingestelde waarde te bevestigen.
Faithfulness	1
Answer relevance	0,963394
Answer correctness	0,766729
Beoordeling expert	Ruim voldoende

Bijlage D: schematische weergave van uitgebreid proces

In deze bijlage staat een uitgebreidere versie van het proces in paragraaf 4.3.4 (Zie figuur 1D voor een schematische weergave). In deze uitgebreide versie wordt de informatie uit de data-analyse gecombineerd met de informatie uit de handleidingen. In acties B1 en B2 is een extra stap toegevoegd waarbij het antwoord op basis van data-analyse wordt samengevat (Zie figuur 2D en 3D voor een schematische weergave). Met deze samenvatting wordt gezocht in de vectordatabase naar relevante tekst uit de handleidingen. Deze relevante tekst wordt samen met het oorspronkelijke antwoord toegevoegd aan een nieuwe prompt. In deze prompt wordt gevraagd of de informatie uit de handleidingen gecombineerd kan worden met het oorspronkelijke antwoord. Vervolgens wordt in beide acties gevraagd om het beste antwoord te kiezen.

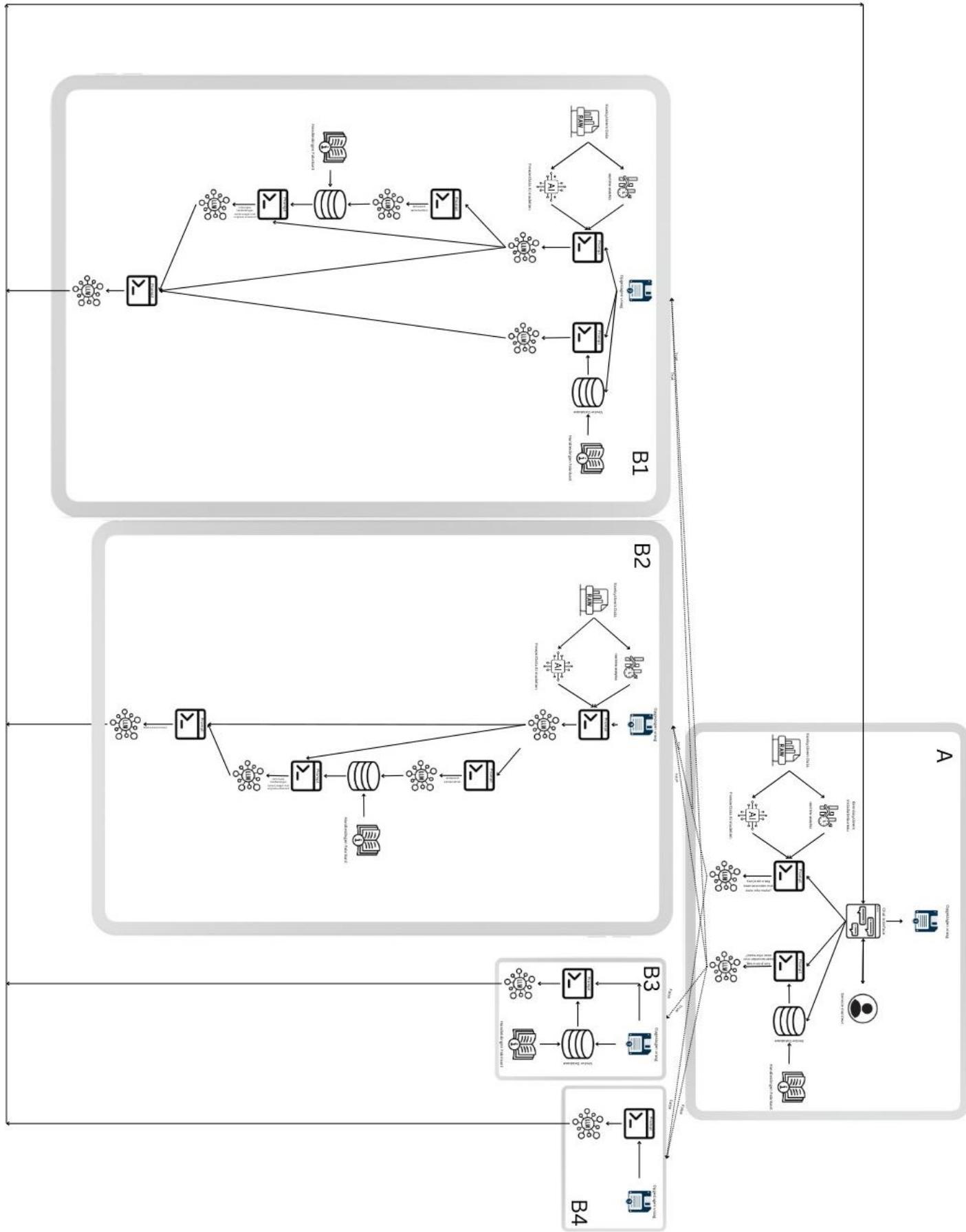
In actie B1 is er een keuze uit drie antwoorden:

- Een antwoord gebaseerd op de data-analyse.
- Een antwoord gebaseerd op de handleidingen.
- Een antwoord gebaseerd op de data-analyse en de handleidingen.

In actie B2 is er een keuze uit 2 antwoorden:

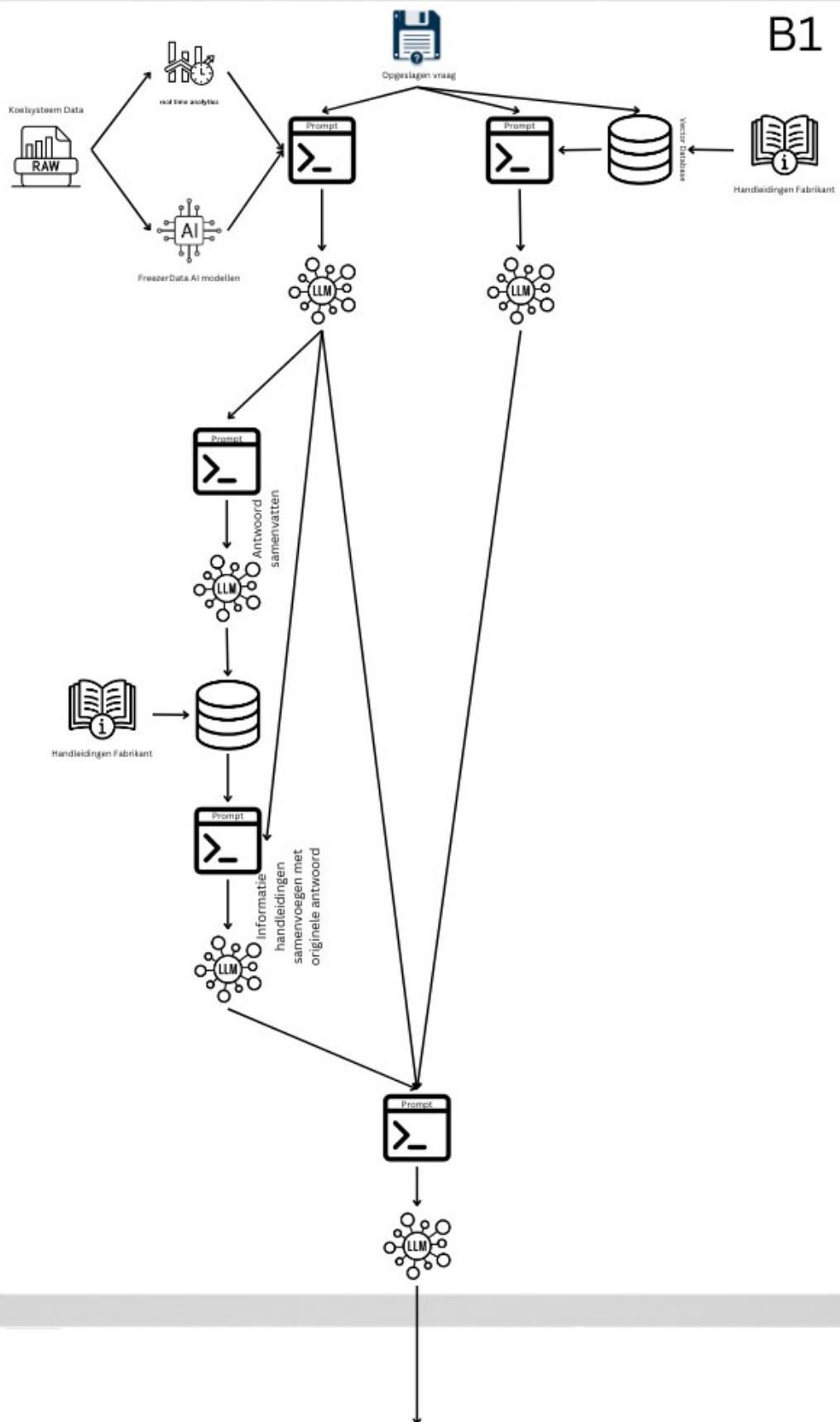
- Een antwoord gebaseerd op de data-analyse
- Een antwoord gebaseerd op de data-analyse en de handleidingen.

Deze oplossing resulteert in minder algemene antwoorden van het model. Echter zorgt deze oplossing ervoor dat er in sommige gevallen informatie uit de handleiding in het antwoord van het model zit dat niet klopte bij de huidige staat van het koelsysteem. Dit probleem zou mogelijk opgelost kunnen worden door verbeteringen aan de prompt of de retriever door te voeren.



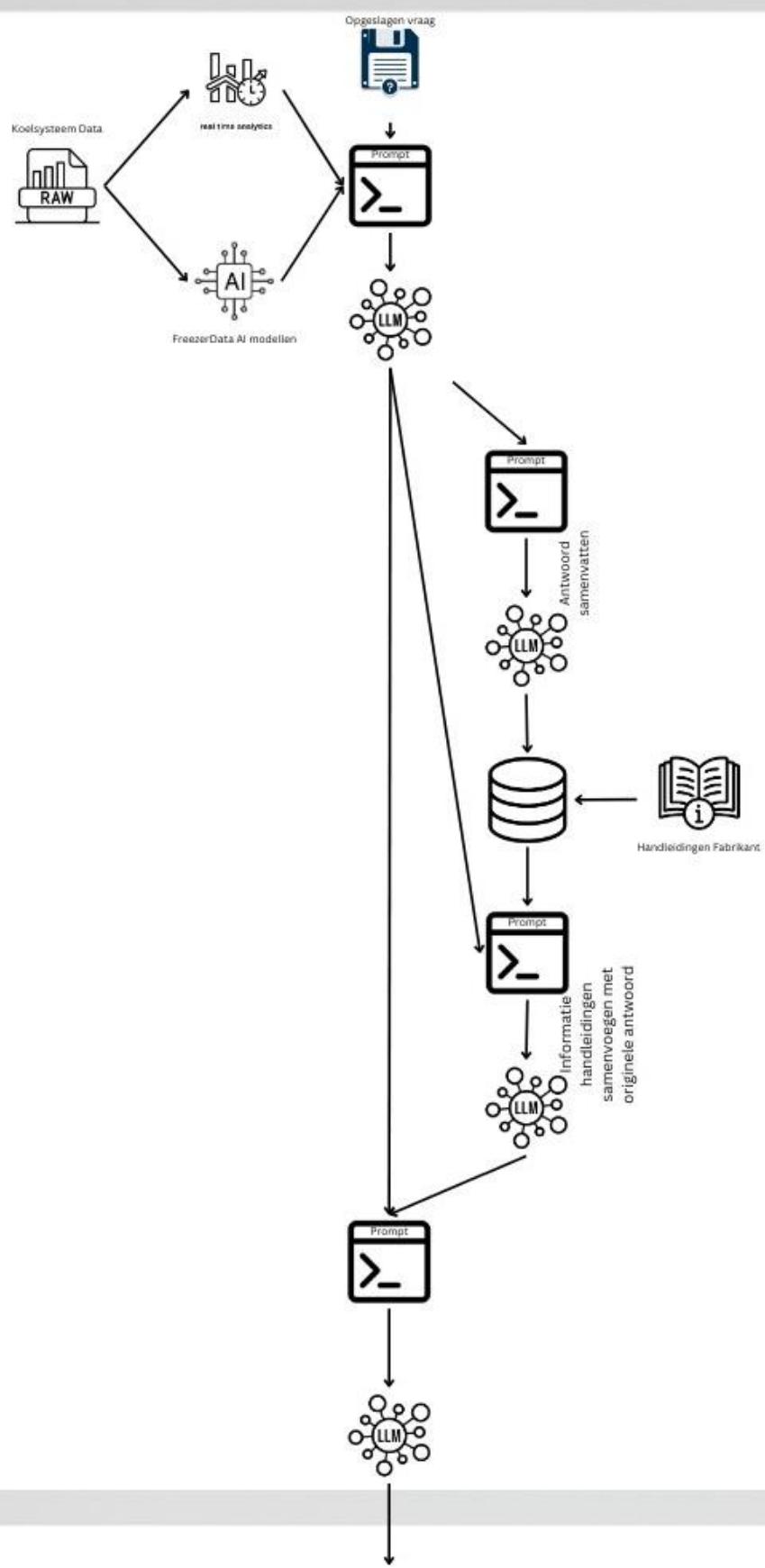
Figuur 1D: Schematische weergave van een uitgebreidere versie van het proces.

B1



Figuur 2D: schematische weergave van uitgebreide actie B1.

B2



Figuur 3D: schematische weergave van uitgebreide actie B2.