PILOTPROSJEKT

# Bruk av kunstig intelligens for reduksjon av fare og ulykkesrisiko grunnet fallende konstruksjoner

Havindustritilsynet

**Rapportnr.:** 2024-2402, Rev. 1

**Dokumentnr.:** 2447589

**Dato:** 2024-12-13

| | | |
|---|---|---|
| Prosjektnavn: | Pilotprosjekt | DNV AS Digital Solutions |
| Rapporttittel: | Bruk av kunstig intelligens for reduksjon av fare og ulykkesrisiko grunnet fallende konstruksjoner | Veritasveien Høvik 1363 |
| Oppdragsgiver: | Havindustritilsynet, Professor Olav Hanssens vei 10 4021 STAVANGER Norway | Norway<br>Tel: 65759900<br>945 748 931 |
| Kontaktperson: | Morten Langøy | |
| Dato: | 2024-12-13 | |
| Prosjektnr.: | 10528122 | |
| Org. enhet: | Performance and Assurance Solutions -4100-NO | |
| Rapportnr.: | 2024-2402, Rev. 1 | |
| Dokumentnr.: | 2447589 | |

Levering av denne rapporten er underlagt bestemmelsene i relevant(e) kontrakt(er):

Avrop nr: A06868 – 01 – 2024 – saksnr. 2024/951

Kort oppdragsbeskrivelse:

Mulighetsstudie for å undersøke hvilken verdi moderne søkemetoder, med bruk av ontologier (kunnskaps-modeller), AI og tekstanalyse, kan gi for å etablere innsikt i trender og læring for fallende konstruksjonselementer. Samt verdi av metodene for andre hendelser.

| Utført av: | Verifisert av: | Godkjent av: |
|---|---|---|
| Digitally signed by Watson, David Ross Date: 2025.02.07 11:12:58 +01'00' | Digitally signed by Søfferud, Mario Date: 2025.02.07 13:35:37 +01'00' | Myrvang, Petter Digitally signed by Myrvang, Petter Date: 2025.02.07 12:26:58 +01'00' |
| David Watson<br>Senior Innovation Lead | Mario Søfferud<br>Principal Engineer | Petter Myrvang<br>Senior Principle Innovation Lead |
| Suyuthi, Abdillah Digitally signed by Suyuthi, Abdillah Date: 2025.02.10 12:11:24 +01'00' | | |
| Abdillah Suyuthi<br>Principle Data Scientist | | |
| Aneeq Ahsan<br>Digital Transformation Consultant | | |

Internt i DNV er informasjonen i dette dokumentet klassifisert som:

Kan dokumentet bli distribuert internt i DNV etter en gitt dato?

Nei Ja

☒ Open -- --

☐ DNV Restricted

☐ DNV Confidential ☐ ☐

☐ DNV Secret

Flere personer som er autorisert til å distribuere dette dokumentet internt i DNV:

Keywords AI, KI, søk, ontologi, LMM, maskinlæring, språkmodeller, kunnskapsgrafer

| Rev. no. | Date | Reason for issue | Prepared by | Verified by | Approved by |
|---|---|---|---|---|---|
| 1 | 2024-12-13 | Minor edits | DAVWAT<br>SUYUT<br>ANEAHS<br>MINANG<br>SIRIA | MARSO | PMYR |

# UAVHENGIGHET, UPARTISKHET OG BEGRENSNINGER I RÅDGIVNINGENS UTSTREKNING

Dette dokumentet inneholder innhold levert av DNV. Vær oppmerksom på følgende:

**Etiske uavhengighetstiltak**
For å opprettholde den nødvendige integritet og upartiskhet som er essensielt for våre tredjepartsroller knyttet til samsvarsvurderinger, utfører DNV innledende interessekonfliktvurderinger før vi påtar oss engasjement i tilknytning til rådgivningstjenester.

**Rolleprioritet**
Denne rapporten er utarbeidet av DNV i sin rådgivende kapasitet, etter at vi har gjort interessekonfliktvurderinger. Innholdet i rapporten er adskilt fra DNVs ulike roller som uavhengig leverandør av tredjeparts tjenester knyttet til samsvarsvurdering. Hvor overlapp eksisterer mellom disse to typene av tjenester, vil tredjeparts tjenester knyttet til samsvarsvurdering utført av DNV være uavhengige av rådgivning som er gitt på vegne av DNV og de vil ha forrang over de rådgivende tjenestene som ytes.

**Fremtidige tredjeparts tjenester knyttet til samsvarsvurdering**
Innholdet i dette dokumentet vil ikke forplikte eller påvirke DNVs uavhengige og upartiske dømmekraft eller utfallet i eventuelle fremtidige tredjeparts tjenester knyttet til samsvarsvurdering som utføres av DNV hvor det kan være en viss tilknytning og sammenheng mellom rådgivingen som er gjort og den fremtidige tredjeparts tjenesten knyttet til samsvarsvurdering som skal ytes.

**Gjennomgang av overholdelse**
DNVs overholdelse av etiske regler og bransjestandarder når det gjelder skille av DNVs ulike roller og tjenester er underlagt periodiske eksterne gjennomganger.

# Innholdsfortegnelse

# 1 OPPSUMMERING – EXECUTIVE SUMMARY

Denne rapporten, utarbeidet av DNV for Havindustritilsynet (Havtil), utforsker hvordan kunstig intelligens (KI) kan brukes i arbeidet med å redusere risikoen for ulykker. Rapporten fokuserer på å forbedre læring og innsikt fra eksisterende data ved hjelp av moderne digitale teknologier som KI, språkteknologi og kunnskapsmodeller (ontologier). I pilotstudien er hendelser med fallende konstruksjonselementer (*dropped structural objects*) brukt.

**Bakgrunn:** Havtil har samlet omfattende informasjonsressurser gjennom flere tiår, inkludert rapporter og databaser fra tilsyn, granskinger og skadeanalyser. Disse dataene er av høy kvalitet, men vanskelig tilgjengelige for systematisk læring. Denne rapporten identifiserer behovet for å utnytte ny teknologi for å gjøre denne informasjonen mer tilgjengelig og nyttig.

**Metode:** Rapporten beskriver en metodisk tilnærming som inkluderer:

- Bruk av ekspertise innen skadeanalyse, gransking, risikostyring, KI, språkteknologi og ontologier.
- Utvikling av kunnskapsmodeller for å strukturere data.
- Implementering av LLM-agenter (Large Language Models) for å forbedre søk og dataanalyse.
- Testing og evaluering av prototyper.

**Resultater:** Rapporten fremhever flere nøkkelresultater:

- Identifisering og evaluering av testscenarier.
- Utvikling av en arkitektur for KI-baserte søkeløsninger.
- Opprettelse av kunnskapsgrafer for å strukturere og integrere data.
- Forbedring av databruk og søkbarhet gjennom bruk av ontologier.
- Initielle resultater med å strukturere ustrukturert informasjon.

**Anbefalinger:** Rapporten gir flere anbefalinger for videre arbeid:

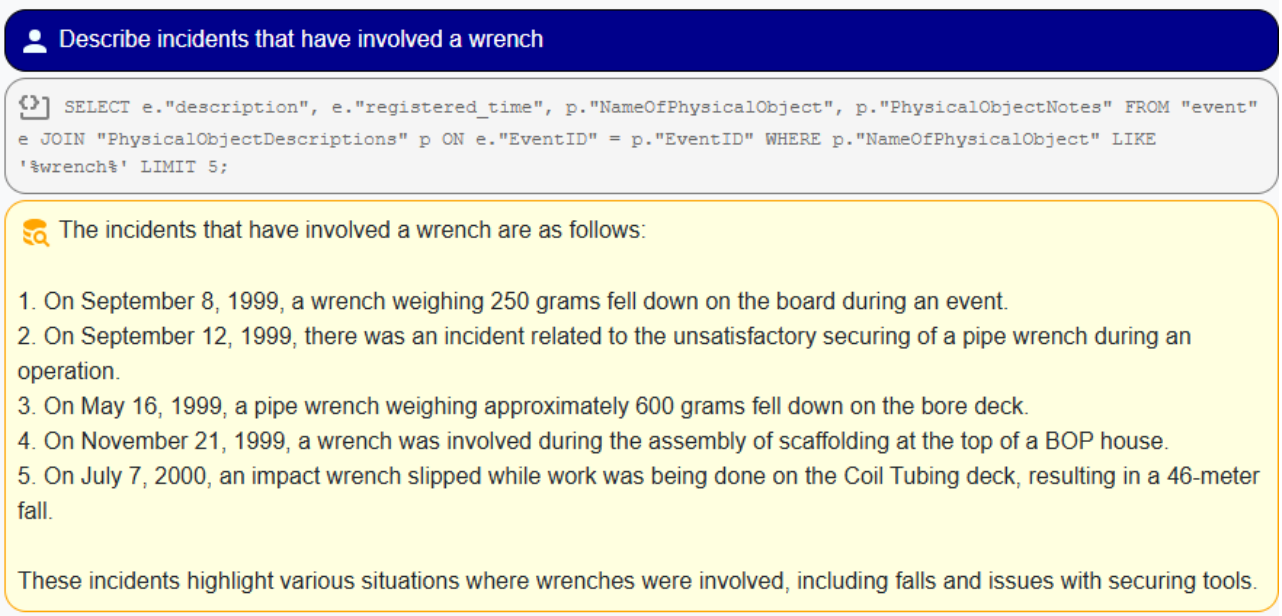
- Utvidelse av datakilder for å inkludere granskningsrapporter og operatørdata.
- Automatisering av evaluering av LLM-resultater.
- Samarbeid med industrien for å dele læringsdata og forbedre sikkerheten på tvers av operatører.

**Konklusjon:** Integrering av KI og moderne analyseteknikker har stort potensial for å transformere måten vi lærer av hendelser på, og sikre at kunnskap blir mer tilgjengelig og anvendelig. Dette vil bidra til bedre risikostyring og sikkerhet i offshore-industrien.

Rapporten understreker viktigheten av samarbeid mellom Havtil, operatører og teknologieksperter for å realisere disse målene og forbedre sikkerheten på norsk sokkel.

**Prototype:** DNV har også investert ekstra tid i dette prosjektet gjennom sitt innovasjonsbudsjett og har utviklet en fungerende prototype i form av en webapplikasjon. Denne webapplikasjonen har gjort testingen mye raskere sammenlignet med Python-notebook-versjonen. Resultatene viser at verktøyet har oppnådd flere viktige milepæler. For det første har det eliminert hallusinasjoner, noe som betyr at verktøyet ikke finner opp falske resultater når det mangler data. For det andre kan verktøyet finne informasjon i fritekst som ikke er mulig å oppdage med nåværende metoder, som for eksempel Power BI-filtrering.

Denne utviklingen representerer et betydelig fremskritt i prosjektet, og demonstrerer potensialet for å forbedre datainnsamling og analyse på måter som tidligere ikke var mulig. Ved å bruke avanserte teknologier og innovative tilnærminger, viser DNV sitt engasjement for å drive fremtidens løsninger innen databehandling og analyse.




*Figur 1: Skjermbilde av et eksempelspørsmål og svaret.*

# 2 BAKGRUNN

Havindustritilsynet (Havtil, tidligere Ptil) har over flere tiår samlet omfattende informasjonsressurser gjennom sin aktivitet på norsk sokkel. Disse ressursene, som inkluderer rapporter og databaser fra tilsyn, granskinger, skade- og årsaksanalyser, samt oppfølging av hendelser, holder høy kvalitet, men er ikke lett tilgjengelige for systematisk læring og innsiktsarbeid. Utfordringen ligger i at mye av informasjonen er lagret i formater som gjør det krevende å trekke ut læring på en effektiv måte. Rapportene er ofte i PDF-format, som krever manuell gjennomgang av kvalifisert personell, noe som begrenser omfanget av læring som kan oppnås.

For å møte denne utfordringen ser Havtil et stort potensial i å utnytte ny digital teknologi, spesielt kunstig intelligens (KI), språkteknologi og kunnskapsmodeller (ontologier). Ved å kombinere disse teknologiene med faglig ekspertise, kan man gjøre læring mer tilgjengelig, muliggjøre dypere innsikt i kritiske problemstillinger og avdekke nye sammenhenger både innenfor og på tvers av operatører, fagområder og installasjoner.

I dag pågår det flere initiativer hos Havtil med mål om å forbedre oppfølging og læring fra hendelser, samt forbedre granskingsmetodikk og risikorapportering (RNNP). Selv om rapporter fra hendelser, tilsyn og granskinger er tilgjengelige på Havtils nettsider, er funksjonaliteten for søk og tilgjengelighet begrenset. Tidligere studier og piloter på området har vist at det er mulig å forbedre læringsutbyttet, men resultatene har vært begrenset grunnet umoden teknologi og variabel datakvalitet.

I løpet av to pågående prosjekter for Havtil har DNV gjennomført flere intervjuer som har gitt oss dyp innsikt i Havtils arbeidsprosesser og utfordringer. Disse prosjektene, som inkluderer utviklingen av en fremtidig hendelsesdatabase «Oppfølging og læring av hendelser» og forbedring av granskingsmetoder «Videreutvikling av granskninger», har identifisert flere områder der vår foreslåtte løsning kan gi betydelige gevinster.

For eksempel har det blitt tydelig at det er behov for å tolke informasjon som ofte ligger “mellom linjene” i eksisterende rapporter. Dette krever i dag manuelle, kvalitative analyser, da de bakenforliggende årsakene til hendelser sjelden er direkte beskrevet. Havtils granskningsteam har også uttrykt et behov for bedre tilgang til informasjon fra tidligere lignende granskinger. Dette er spesielt viktig når spesifikke tekniske detaljer må hentes ut fra ulike dokumenter og systemer.

Ledelsen i Havtil har også pekt på utfordringer med systematisk sammenligning av tilsynseffekter og vurdering av aktørers overholdelse av regelverket. Det har vært vanskelig å finne en enhetlig metodikk for dette arbeidet. Videre har vi sett et behov for analyser på tvers av granskinger for å få et bedre bilde av utfordringene i næringen. Det er et spesielt behov for å samarbeide med industrien for å bruke deres datagrunnlag, da Havtils egne granskinger kan være for få til å gi en tilstrekkelig bred innsikt. Dette er noe vi ønsker å adressere i vårt pågående interne innovasjonsprosjekt i samarbeid med operatørene. Vi har som mål å integrere dette i senere faser, slik det er beskrevet i vår forstudie.

Vår foreslåtte løsning, som integrerer kunstig intelligens og moderne analyseteknikker, har potensial for å møte disse behovene ved å effektivisere søk, tolkning og sammenstilling av data. Dette vil ikke bare forenkle granskingsarbeidet, men også bidra til å forbedre design- og vedlikeholds prosesser, samtidig som det gir bedre innsikt i utviklingstrekk på tvers av selskapene og bransjen som helhet. Kort sagt, løsningen kan være med å transformere måten vi lærer av hendelser på, og sikre at kunnskap blir mer tilgjengelig og anvendelig i fremtiden. På denne måten kan ny teknologi og metode være med å bidra til et bedre grunnlag for innsikt og læring som igjen vil være viktig for arbeidet med god risikostyring på norsk sokkel.

Med utgangspunkt i ovennevnte utfordringer og muligheter er målet med dette oppdraget å gjennomføre en pilotstudie innen et konkret område: fallende konstruksjons-elementer.

## 2.1 Definisjoner rundt kunstig intelligens

### 2.1.1 Kunstig intelligens (KI)

KI er maskiners evne til å etterligne menneskelig intelligens. Dette inkluderer oppgaver som persepsjon, læring, resonnering og beslutningstaking. KI omfatter felt som maskinlæring, språkprosessering og robotikk.

Eksempler:

- Selvkjørende biler
- Stemmesystemer (f.eks. Siri, Alexa)
- Bildegjenkjenning

### 2.1.2 Maskinlæring (ML)

En gren av KI der maskiner lærer av data uten eksplisitt programmering. ML identifiserer mønstre i data og forbedres med mer trening.

Hovedtyper:

- Supervised learning: Læring med merkede data (f.eks. prisprognoser).
- Unsupervised learning: Oppdager mønstre i umerkede data (f.eks. kundesegmentering).
- Reinforcement learning: Lærer ved prøving og feiling.

Eksempler:

- Svindeldeteksjon
- Anbefalingssystemer (f.eks. Netflix)
- Prediktivt vedlikehold

### 2.1.3 Store språkmodeller (LLM-er)

Avanserte ML-modeller som trener på store mengder tekst for å forstå og generere menneskelignende språk. Bruker teknologier som transformere (f.eks. GPT, BERT).

Muligheter:

- Tekstgenerering
- Sammendrag
- Oversettelse
- Sentimentanalyse

Eksempler:

- GPT (f.eks. ChatGPT)
- BERT

### 2.1.4 Dyp læring

En underkategori av maskinlæring basert på nevrale nettverk med mange lag. Ideelt for komplekse data som bilder og språk.

Eksempler:

- Ansiktsgjenkjenning
- Selvkjørende bilers sensorer
- Avansert oversettelse

### 2.1.5 Andre nøkkelbegreper

Nevrale nettverk: Algoritmer inspirert av hjernen, grunnlaget for dyp læring.

Naturlig språkprosessering (NLP): KI som håndterer menneskelig språk (f.eks. chatboter, tekst-til-tale).

Generativ KI: Skaper nytt innhold som tekst, bilder og video (f.eks. DALL-E, GPT).

Edge KI: Lokalt prosesserende KI på enheter som smarttelefoner og IoT.

## 2.2 En introduksjon til ontologier

Ontologier er strukturerte rammeverk som representerer kunnskap som et sett av konsepter innenfor et domene, og relasjonene mellom disse konseptene. De gir en formell og eksplisitt spesifikasjon av en delt konseptualisering, som muliggjør konsekvent forståelse og kommunikasjon på tvers av systemer og interessenter. En ontologi inkluderer typisk:

1. **Klasser (konsepter):** Enhetene eller typer ting innenfor domenet. For eksempel kan klasser i et hendelsesrapporteringssystem inkludere Hendelse, Sted, Konstruksjonselement og Dato.
2. **Egenskaper (attributter og relasjoner):** Kjennetegn ved klassene og relasjonene mellom dem. For eksempel kan en Hendelse-klasse ha egenskaper som hendelses dato, sted og involverte konstruksjonselementer.
3. **Forekomster:** De spesifikke eksemplene på klasser, som en bestemt hendelse som skjedde på en bestemt dato og sted.
4. **Regler og begrensninger:** De logiske begrensningene og forretningsreglene som gjelder for klassene og egenskapene, som sikrer datakonsistens og gyldighet.

## 2.3 Hvordan ontologier hjelper med databehandling

Ontologier kan vesentlig forbedre lagring, håndtering og bruk av data på flere måter:

1. **Standardisering og konsistens:**
    a. Enhetlig datarepresentasjon: Ontologier gir en standardisert måte å representere data på, og sikrer at all lagret data følger en konsekvent struktur og terminologi. Dette er avgjørende for å integrere data fra forskjellige kilder.
    b. Unngå tvetydighet: Ved å klart definere konsepter og deres relasjoner, eliminerer ontologier tvetydighet, noe som gjør datatolkningen enkel.
2. **Interoperabilitet:**
    a. Kommunikasjon på tvers av systemer: Ontologier gjør det mulig for ulike systemer å forstå og bruke data på en kompatibel måte. Dette er spesielt viktig når data fra forskjellige databaser og formater må integreres.
    b. Semantisk integrasjon: Når data fra forskjellige kilder kombineres, hjelper ontologier med å kartlegge ekvivalente konsepter og relasjoner, noe som sikrer sømløs integrasjon og aggregering.
3. **Forbedret spørring og analyse:**
    a. Semantiske spørringer: Ontologier muliggjør avanserte spørringsmuligheter, slik at brukere kan utføre komplekse spørringer som utnytter relasjonene og hierarkiene definert i ontologien.
    b. Datautledning: Ved å bruke de logiske reglene og begrensningene definert i ontologien, kan systemer utlede ny informasjon fra eksisterende data, noe som beriker og øker nytten av de lagrede dataene.
4. **Fleksibilitet og skalerbarhet:**
    a. Tilpasning til endringer: Ontologier kan utvikle seg over tid etter hvert som nye konsepter og relasjoner identifiseres. Denne fleksibiliteten lar datasystemet tilpasse seg nye krav uten betydelig omstrukturering.

b. Skalerbar kunnskapsrepresentasjon: Ontologier kan representere store og komplekse domener effektivt, noe som gjør dem egnet for omfattende og varierte datasett.

5. **Forbedret datakvalitet:**

   a. Validering og integritet: Ontologier håndhever regler og begrensninger som sikrer datakvalitet, som obligatoriske egenskaper og gyldige relasjoner. Dette hjelper med å opprettholde integriteten og nøyaktigheten av de lagrede dataene.

   b. Feilreduksjon: Ved å gi en klar struktur og retningslinjer for dataregistrering og lagring, reduserer ontologier sannsynligheten for feil og inkonsekvenser.

6. **Forbedret dataintegrasjon:**

   a. Kontekstuell berikelse: Ontologier legger til kontekst til dataene ved å definere hvordan forskjellige informasjonsbiter er relatert, noe som gjør det enklere å kombinere og sammenligne data fra forskjellige kilder.

   b. Ontologi-drevne datasystemer: Systemer designet med ontologier kan automatisk kategorisere og lagre data i riktig format, noe som forenkler databehandling og gjenfinning.

Kort oppsummert, ontologier er kraftige verktøy for strukturering og lagring av data, som sikrer konsistens, forbedrer interoperabilitet, muliggjør avanserte spørringer og analyser, og forbedrer den generelle datakvaliteten. De gir et robust rammeverk for å håndtere komplekse dataomgivelser, noe som gjør dem uunnværlige i moderne datalagrings- og håndteringsløsninger.

## 2.4 Kunnskapsgrafer

Gartner (Jaffri A. 2024) beskriver kunnskapsgrafer som en måte å levere semantisk aktivert databehandling som muliggjør et bredt spekter av KI-applikasjoner. For å lykkes med å bygge kunnskapsgrafer på bedriftsnivå, må data- og analyseledere ta en smidig tilnærming til utviklingen av slike grafer.

Hovedfunn fra Gartner:

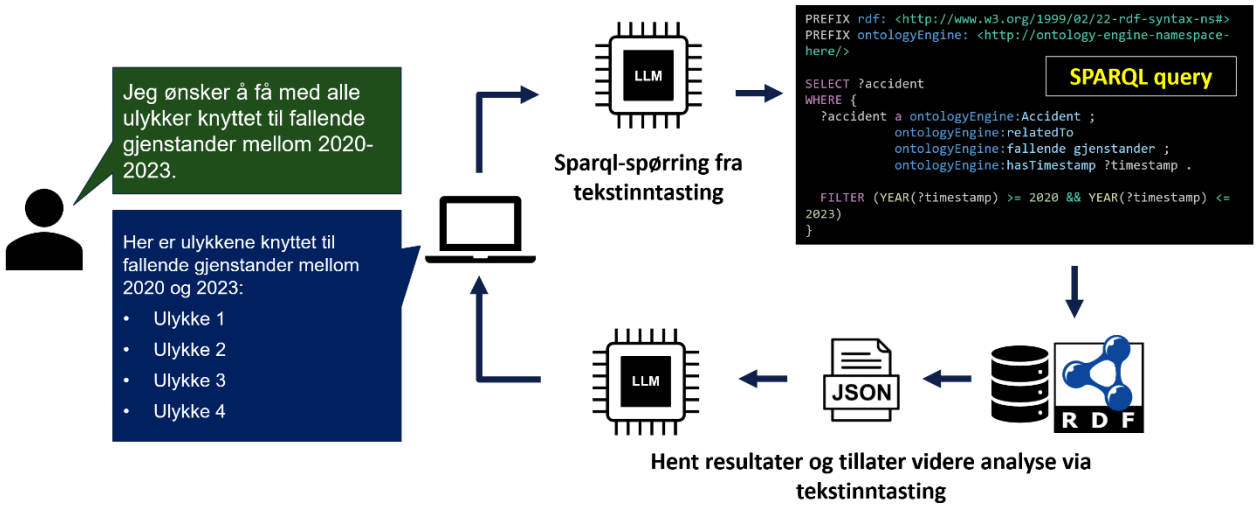
- Den fleksible, modulære og åpne naturen til kunnskapsgrafbasert datalevering gjør det enklere å sikre semantisk konsistens i data på tvers av virksomheten. Dette gjør det mulig for forretningsbrukere, programvareutviklere og dataforskere å finne, forstå og bruke dataene de trenger.
- Kunnskapsgrafer kan fungere som en selvstendig KI-ressurs eller brukes sammen med maskinlæringsmodeller i en sammensatt KI-tilnærming. Dette gir mulighet for kunnskapsoppdagelse, søk og gjenfinning, anbefalinger og plattformer for beslutningsintelligens.

### 2.4.1 Hovedforskjellen mellom kunnskapsgrafer og ontologier

Kunnskapsgrafer bygges ofte ved hjelp av ontologier som skjema, noe som gjør det mulig å strukturere data i et formelt rammeverk. Likevel legger kunnskapsgrafer vekt på praktisk bruk av data, mens ontologier fokuserer på teoretisk organisering av denne kunnskapen.

## 2.5 Bruk av kunstig intelligens ved gjenfinning av innhold

Store språkmodeller (LLM) kan effektivt konvertere en tekstspørring til en SPARQL-spørring for grafdatabaser. SPARQL is a powerful query language and protocol used for retrieving and manipulating data stored in Resource Description Framework (RDF) format. Ved hjelp av naturlig språkbehandling, kan LLM-en analysere og tolke intensjonen bak brukerens spørsmål. Deretter genererer den en tilsvarende SPARQL-spørring som nøyaktig henter den ønskede informasjonen fra grafdatabasen. Resultatene fra SPARQL-spørringen presenteres deretter på en brukervennlig måte, ofte gjennom visualiseringer eller strukturerte rapporter, som gjør det enkelt for brukeren å utføre videre analyse og trekke innsikter. Dette muliggjør en sømløs overgang fra komplekse databaser til lettforståelige data for sluttbrukeren.




*Figur 2: Bruk av kunstig intelligens for å finne og analysere data*

## 2.6 Fordel med ontologibruk fremfor LLM alene

Bruken av ontologier forbedrer søkeresultatene fra en database sammenlignet med bruk av kun store språkmodeller (LLM) av flere grunner. For det første gir ontologier en standardisert og konsistent representasjon av data ved å definere klare konsepter og relasjoner mellom dem. Dette eliminerer tvetydighet og sikrer at alle data blir tolket på samme måte, noe som fører til mer presise søkeresultater.

For det andre muliggjør ontologier semantisk integrasjon, hvor data fra forskjellige kilder kan kobles sammen og forstås i kontekst. Dette er spesielt viktig når man søker i komplekse datasett med variert opprinnelse. Ontologier hjelper med å mappe ekvivalente konsepter og relasjoner, noe som sikrer at søkene gir omfattende og relevante resultater.

Til slutt forbedrer ontologier søkbarheten gjennom hierarkisk strukturering av data. Dette gir mulighet for mer granulære søk, hvor brukeren kan spesifisere nøyaktige underkategorier for å finne presis informasjon. Mens LLM-er kan tolke og forstå naturlig språk, gir ontologier den strukturelle rammen som sikrer at søkeresultatene er både nøyaktige og omfattende, noe som gir en mer pålitelig og brukervennlig søkeopplevelse.

## 2.7 Effekt av «state-of-the-art» metode fremfor «manuelle metoder»

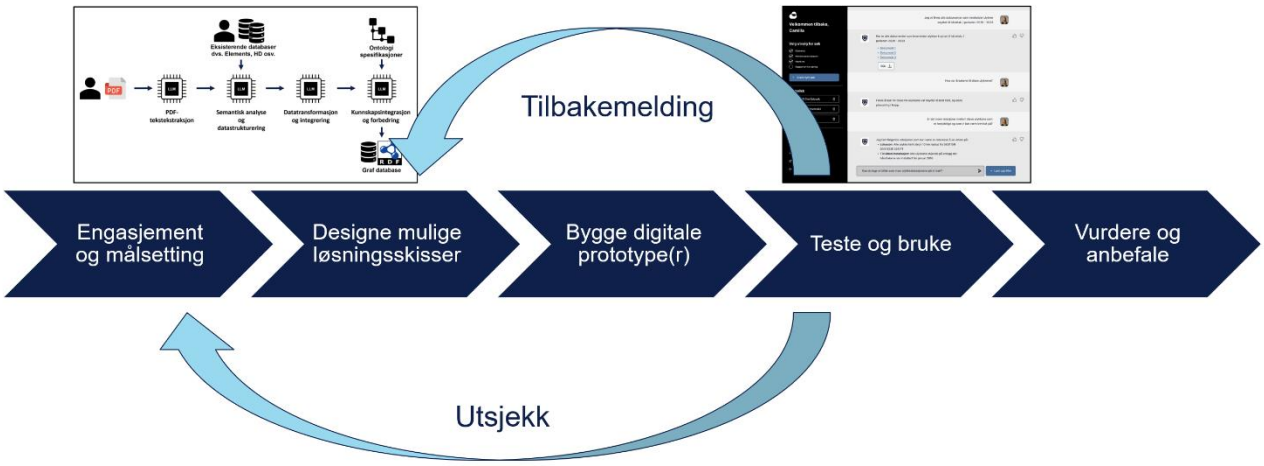
Fordelen med å bruke søkemetoder basert på store språkmodeller (LLM) og ontologier fremfor manuelle metoder for å søke i databaser er betydelig. Først og fremst gir LLM-er og ontologier en mer effektiv og nøyaktig søkeprosess. LLM-er kan tolke naturlig språk, noe som gjør det mulig for brukere å skrive søkespørringer på en intuitiv og enkel måte, uten å måtte kjenne til spesifikke søkesyntakser eller kommandoer.

Ontologier bidrar til å strukturere og standardisere dataene, noe som eliminerer tvetydighet og sikrer at alle søk blir konsistente og presise. Dette er spesielt nyttig når man arbeider med store og komplekse datasett, hvor manuelle søk kan være tidkrevende og utsatt for feil.

I tillegg muliggjør kombinasjonen av LLM-er og ontologier semantisk søk, hvor systemet forstår og tolker meningen bak spørringene, og dermed gir mer relevante og omfattende resultater. Dette reduserer behovet for flere manuelle søk og filtrering av irrelevante resultater, noe som sparer tid og ressurser.

Til slutt, ved å bruke LLM-er og ontologier, kan data fra forskjellige kilder integreres sømløst, noe som gir en helhetlig oversikt og dypere innsikt. Dette er ofte vanskelig å oppnå med manuelle søkemetoder som kan kreve mye arbeid for å samle og tolke data fra ulike systemer.

# 3 METODEBESKRIVELSE




*Figur 2: Skisse av metodebeskrivelse*

DNV startet dette prosjektet med et sterkt engasjement og klare målsettinger for å sikre at vi oppnådde de ønskede resultatene. Dette ble gjort gjennom en workshop for avklaring av omfang. Omfanget for denne pilotstudien var avgrenset til fallende konstruksjons-elementer. Vi designet mulige løsningsskisser som svarte på de identifiserte behovene og kravene med utgangspunkt i dette området. Dette innebar å utvikle flere konsepter og evaluere potensialet for å løse de spesifikke utfordringene vi sto overfor.

Deretter bygget vi digitale prototyper basert på løsningsskissene beskrevet i mulighetsstudiet. Disse prototypene ble testet og brukt i bruksscenarier for å vurdere deres effektivitet og brukervennlighet. Case-studiene vi brukte for å teste det nye digitale søkeverktøyet, var basert på innsikten fra en digital workshop med Havtil. I denne workshopen samlet vi informasjon om hvilke typer data og spørsmål saksbehandlere og granskningsledere ofte søkte etter i sitt daglige arbeid. Dette hjalp oss med å utvikle realistiske og relevante case-studier som reflekterte de faktiske behovene og utfordringene de møtte.

Ved å bruke disse case-studiene kunne vi teste søkeverktøyets evne til å håndtere komplekse spørringer og levere nøyaktige og nyttige resultater. En viktig del av denne prosessen var den iterative syklusen med utsjekk og tilbakemelding. Etter hver testfase gjennomførte vi en grundig evaluering av resultatene og innhentet tilbakemeldinger fra DNVs fageksperter. Disse tilbakemeldingene ble brukt til å justere og forbedre prototypene, slik at vi kontinuerlig kunne optimalisere løsningene.

Til slutt vurderte vi resultatene og kom med anbefalinger for videre arbeid. Dette inkluderte å evaluere metodene og teknologiene som ble brukt, samt å foreslå forbedringer og justeringer basert på de innsamlede dataene og tilbakemeldingene. Målet er å vise Havtil et «proof of concept» som demonstrerer at denne teknologien kan brukes effektivt og at den har potensial til videre utvikling, dersom Havtil ønsker å fortsette samarbeidet.

## 3.1 Bruk av ekspertise

DNV stilte til rådighet en kombinasjon av ekspertise innen flere områder for dette prosjektet. Ledende digital ekspertise innen kunstig intelligens, språkteknologi og kunnskapsmodellering (ontologier) jobbet tett sammen med fageksperter innen fallende konstruksjons-elementer, materialteknologi, feilanalyser, gransking, rot-årsaksanalyse (RCA), sikring, risikostyring og sikkerhet (HMS). Vår integrerte kompetanse innen innovasjonsmetoder var også en viktig komponent i prosjektet. DNV sine fageksperter arbeidet sammen med Havtil for å forstå hvilke spørsmål de ønsket besvart gjennom den nye søkemetodikken. Denne innsikten var avgjørende for å identifisere nøkkelparametere som var viktige for utviklingen av ontologier og KI-baserte søkeløsninger.

I tillegg hadde flere i teamet god innsikt i og erfaring med Havtil sin virksomhet gjennom prosjekter og engasjement som også var relevante for dette oppdraget.

## 3.2 Utvikling av kunnskapsmodeller

Kunnskapsmodeller eller ontologier vil definere fagområder på en måte som vil hjelpe KI og språkprosessering med å treffe på behovene til løsningen. En slik kombinert tilnærming vil gi bedre presisjon i spørringer, søk og prosessering utfra den gitte sammenhengen det er snakk om.

Ontologieksperter hos DNV vil sammen med fageksperter utarbeide en struktur som legger til rette for semantiske forståelse på tvers av systemer. Det betyr at brukerne kan utføre komplekse søk som utnytter relasjoner og hierarkier definert i ontologien. Dette vil gjøre det lettere å finne relevant informasjon og å trekke innsiktsfulle konklusjoner. Kunnskapsmodeller vil gjøre det mulig for ulike systemer å forstå og bruke data på en kompatibel måte. Dette er spesielt kritisk når data fra forskjellige databaser og formater må integreres for helhetlig analyse.

## 3.3 Bruk av KI og «LLM-agenter»

DNVs KI-eksperter vil kombinere flere lag med "LLM-agenter" (Large Language Models) med kunnskapsmodeller for å skape en kraftig søkeplattform. Denne løsningen vil bli implementert i en ny grafdatabase, som i denne piloten vil være begrenset til et avtalt antall dokumenter relatert til fallende konstruksjonselementer. Søkeresultatene fra denne databasen vil bli evaluert av DNVs fageksperter, og nødvendige iterasjoner av søkemetodikken eller ontologiene vil bli utført for å forbedre nøyaktigheten og relevansen av svarene.

For formålet med smart søk i hendelsesdatabasen (HD) benyttes ulike LLM-agenter til å:

1. Skille mellom småprat og meningsfulle spørsmål og svare passende for hver kategori.
    a. Hvis det er småprat, vil brukeren høflig bli bedt om å stille spørsmål relatert til temaet for databasen. Videre vil det gi eksempler på spørsmål som kan stilles og få svar på. Det vil også gi forbudte forespørsler/spørsmål relatert til noen grunnleggende databasekommandoer.
    b. Hvis det er et meningsfullt spørsmål, vil det vurdere om spørsmålet har tilstrekkelig informasjon til å utføre en forespørsel i databasen. Hvis det ikke er tilstrekkelig, vil det be brukeren om å stille et mer detaljert spørsmål.
2. Forstå spørsmålet som brukeren stiller. Det betyr hvilken type informasjon eller data brukeren leter etter.
3. Formulere SQL-forespørsel.
4. Formulere svaret basert på spørsmålet, SQL-forespørselen og forespørselsresultatene.

For formålet med forståelse av PDF-dokumenter benyttes LLM-agenten til å:

1. Produsere innebyggingsvektorer av tekstbiten.
2. Produsere innebyggingsvektor av spørsmålet.
3. Formulere svaret basert på spørsmålet, relaterte tekstbiter og deres metadata.

## 3.4 Vurdering av prototype

Under prosjektet evaluerte DNVs fageksperter resultatene fra modellen og ga tilbakemeldinger til designet i en iterativ prosess. DNV har demonstrert den første fungerende prototypen på et møte med Havtil før prosjektlevering. Prototypen kan gjøres tilgjengelig for samtesting for medlemmer av Havtil dersom Havtil ser verdi i å fortsette utviklingen av dette verktøyet. Når kjerneverktøyet er nærmere et implementeringsnivå,

kan ytterligere funksjonalitet legges til slik at verktøyet blir en nyttig komponent i Havtils arbeidsflyt rundt hendelser.

Det er viktig å merke seg at noen av spørsmålene Havtil ønsker å stille KI-verktøyet ikke kan besvares med det nåværende datasettet. For eksempel kan ikke spørsmål om årsakene til hendelser besvares godt, da denne informasjonen sjelden er registrert i Hendelsesdatabasen. For å kunne besvare slike spørsmål må Hendelsesdatabasen suppleres med data fra hendelsesrapporter fra enten operatøren eller Havtil, eller alternativt fra operatørenes hendelseshåndteringssystemer, hvor mer omfattende data er tilgjengelig.

### 3.4.1 Verifisering av resultater

Det finnes flere måter å verifisere resultatet av det smarte søket på.

1. Vise den produserte SQL-forespørselen til brukeren for å la brukeren vurdere relevansen av den produserte SQL-forespørselen i forhold til spørsmålet. Denne tilnærmingen krever en ekspertbruker for å utføre verifikasjonen.
2. Viser trinnvise prosesser bak kulissene slik at (ekspert)brukeren kan se om det gir mening.
3. Beregne vurderingsmetrikker for å tillate automatisk resultatverifisering (Es et al. 2023), som:
    a. Troverdighet, som måler hvor godt svaret er forankret i den gitte konteksten.
    b. Svarrelevans, som måler hvor godt det genererte svaret adresserer det faktiske spørsmålet som ble stilt.
    c. Kontekstrelevans, som måler graden av fokus i den hentede konteksten, slik at den inneholder så lite irrelevant informasjon som mulig.

Forutsatt at kriteriene for akseptabelt eller ikke akseptabelt er bestemt, krever denne tilnærmingen ikke en ekspertbruker for å utføre verifikasjonen. Den kan være fullstendig automatisert, og resultatet kan vises, for eksempel som et trafikklys i brukergrensesnittet. Det er ønskelig at verktøyet også skal kunne referere til kildedokumentene som ble brukt for å konstruere svaret. Dette kan for eksempel være relevante saksnumre slik at brukeren kan manuelt sjekke dataene om ønskelig.

#### 3.4.1.1 Kunnskapsintegrasjon og forbedring

Kunnskapen som ble utvunnet ved bruk av LLM og kunnskapsgrafer er i strukturert format, og i denne første prototypen ble den utvunnede kunnskapen verifisert manuelt for et tilfeldig utvalg av hendelsesdata. Integrasjonen innebar ganske enkelt å samle all data på ett sted med forhåndsdefinert skjema slik at LLM forstår datastrukturen bedre. Det var viktig å gi beskrivelser av alle datafelt slik at en LLM forstår hvor den skal finne hva når en bruker spør etter det. Det anbefales imidlertid at informasjon som utvinnes ved bruk av kunnskapsgraf og LLM i fremtiden verifiseres automatisk eller semi-automatisk på hele datasettet for å sikre høyere kvalitet på utvinningen.

## 3.5 Data

Datakilden for prosjektet ble bestemt som en del av det innledende oppstartsmøtet. En Excel-utskrift av Hendelsesdatabasen med 17 815 unike oppføringer, hvorav 3 819 involverte fallende konstruksjons-elementer. Det ble også avtalt å sende over ytterligere PDF-rapporter fra både Havtil og noen eksempler fra bransjen, totalt 12 rapporter. Ytterligere 3 rapporter skrevet på engelsk ble senere hentet fra Havtil.no ved hjelp av søkefunksjonen, noe som gjorde det mulig for DNVs KI-ekspert å raskt kvalitetskontrollere resultatet av PDF-tekstekstraksjonen på et språk de er mer kjent med.

Det avtalte antallet dokumenter og data gjennomgikk kvalitetskontroll før de ble lagret i en midlertidig database som en del av pilotprosjektet. DNV er innforstått med at Havtil for tiden undersøker alternative hendelseshåndteringssystemer, noe som kan påvirke hvordan data vil bli håndtert i fremtiden.

### 3.5.1 Hvordan de forskjellige datakilder har blitt brukt

#### 3.5.1.1 Hendelsesdatabasen

Hendelsesdatabasen brukes som hovedkilde for det smarte søket. Kolonnenavnene i HD må beskrives tydelig for at LLM skal kunne arbeide med dem. Under operasjonen tar LLM hensyn til datastrukturen og beskrivelsene av kolonnene, men ikke selve dataene. Dataene som hentes av LLM presenteres for brukeren i et brukervennlig format.

#### 3.5.1.2 PDF-dokumenter

PDF-rapporter, hvis de er tilgjengelige, kan brukes som tilleggsinformasjon for hver post i databasen. Informasjonen fra PDF-rapporten vil bli ekstrahert og lagt til de ulike eksisterende kolonnefeltene, hvis mulig. Alternativt kan nye kolonner eller en helt ny tabell koblet til hovedhendelsesdatabasen opprettes.

### 3.5.2 Beste praksis innen digitalisering

Fordelen med at DNV bygger en løsning basert på store språkmodeller (LLM) og ontologier for Havtil, er at det sikrer overholdelse av DNVs anbefalte praksis for digital tillit. Ved å integrere LLM og ontologier, kan DNV levere en robust og pålitelig plattform som forbedrer nøyaktigheten og relevansen av datahåndtering og søk. Dette fremmer transparens og integritet i dataene, noe som er i tråd med DNVs standarder for digital sikkerhet og pålitelighet. Denne tilnærmingen bidrar til økt effektivitet og kvalitet i operasjonene, og styrker samtidig tilliten til digitale løsninger innen offshore-industrien.

### 3.5.3 Datakvaliteten

Løsningen i seg selv vil ikke forbedre datakvaliteten i den eksisterende databasen. For å få en pålitelig løsning trenger vi data av høy kvalitet som utgangspunkt. Derfor vil enhver mulighet til å kontrollere og øke kvaliteten på dataene bli utnyttet under prosessen. Videre kan denne innsatsen brukes som forslag til forbedring av fremtidig innsamling og håndtering av hendelsesdata av Havtil.

Ved å bruke LLM-er for å tolke og strukturere data, og ontologier for å standardisere terminologi og relasjoner, minimiseres feil og tvetydigheter. Dette gir en enhetlig og pålitelig datainfrastruktur hvor informasjonen er lett tilgjengelig og søkbar. Forbedret datakvalitet betyr at beslutninger kan tas på grunnlag av mer presise og komplette data, noe som styrker den overordnede effektiviteten og påliteligheten i virksomheten.

# 4 RESULTATER

## 4.1 Identifisering og evaluering av testscenarier

Testscenarier ble identifisert basert på vår digitale workshop med Havtil. DNVs fageksperter diskuterte deretter prosessen de ville bruke for å besvare spørsmålene knyttet til testscenariene, og om det i det hele tatt var mulig å besvare spørsmålene basert på datasettet. De etterfølgende delkapitlene viser eksempel på spørsmål innenfor utvalgte tema. Ettersom datagrunnlaget blir større og modellen mer avansert vil det være naturlig å kombinere disse temaene i et eller flere spørsmål, men for å demonstrere modellens kapasitet er temaene avgrenset slik.

### 4.1.1 Årsaker og rotårsaker

Økt innsikt i (direkte) årsaker og rotårsaker fra hendelsesdatabasen (HD) er et åpenbart nyttig og relevant resultat ved bruk av smarte søk. I HD er det eksempelvis 344 treff på «årsak» (og ingen treff på «rotårsak»), men som et resultat av det narrative formatet på beskrivelsen er det i mange tilfeller ikke-relevante treff, for eksempel:

- «Årsak til forsinket varsling (…)» - dette er en gjenganger
- «Årsaken var korrosjon» - forveksling mellom skademekanisme og årsak
- «…(kunne) forårsaket…» - som i «(kunne) føre til»

Tar man med både «granskning» og «årsak» i beskrivelsesfeltet får man 64 treff i HD (av totalt 3818). I mange tilfeller er det kopiert inn oppsummering fra en granskningsrapport og reelle årsaker som er identifisert ligger inne i hendelsesdatabasen. Den lave andelen på under 2 % indikerer likevel at hovedtyngden av opplysninger om årsaker og evt. rotårsaker er å finne i de detaljerte granskningsrapportene.

Eksempel på spørsmål:

- Hva er typiske årsaker knyttet til hendelser med fallende gjenstander i forbindelse med boring og brønn?
- Hva er typiske årsaker knyttet til hendelser med fallende gjenstander fra stillas?

Svar på slike spørsmål vil, i motsetning til å hente ut enklere statistikk, være en stor jobb å finne manuelt, og er en god oppgave for KI.

### 4.1.2 Trender og utvikling

Innenfor gruppen av spørsmål om trender og utvikling ligger det mye behandling av data både med språkmodeller og statistikk. Eksempel på spørsmålstema kan være:

- Trender, utvikling og statistikk for;
    - en spesifikk aktivitet /operasjon
    - en spesifikk hendelsestype,
    - et spesifikt felt,
    - en spesifikk operatør,
    - en spesifikk årstid eller
    - en spesifikk type installasjon
    - osv.
- Utvikling i hendelser fra år til år med høy alvorlighetsgrad som har ført til granskning.

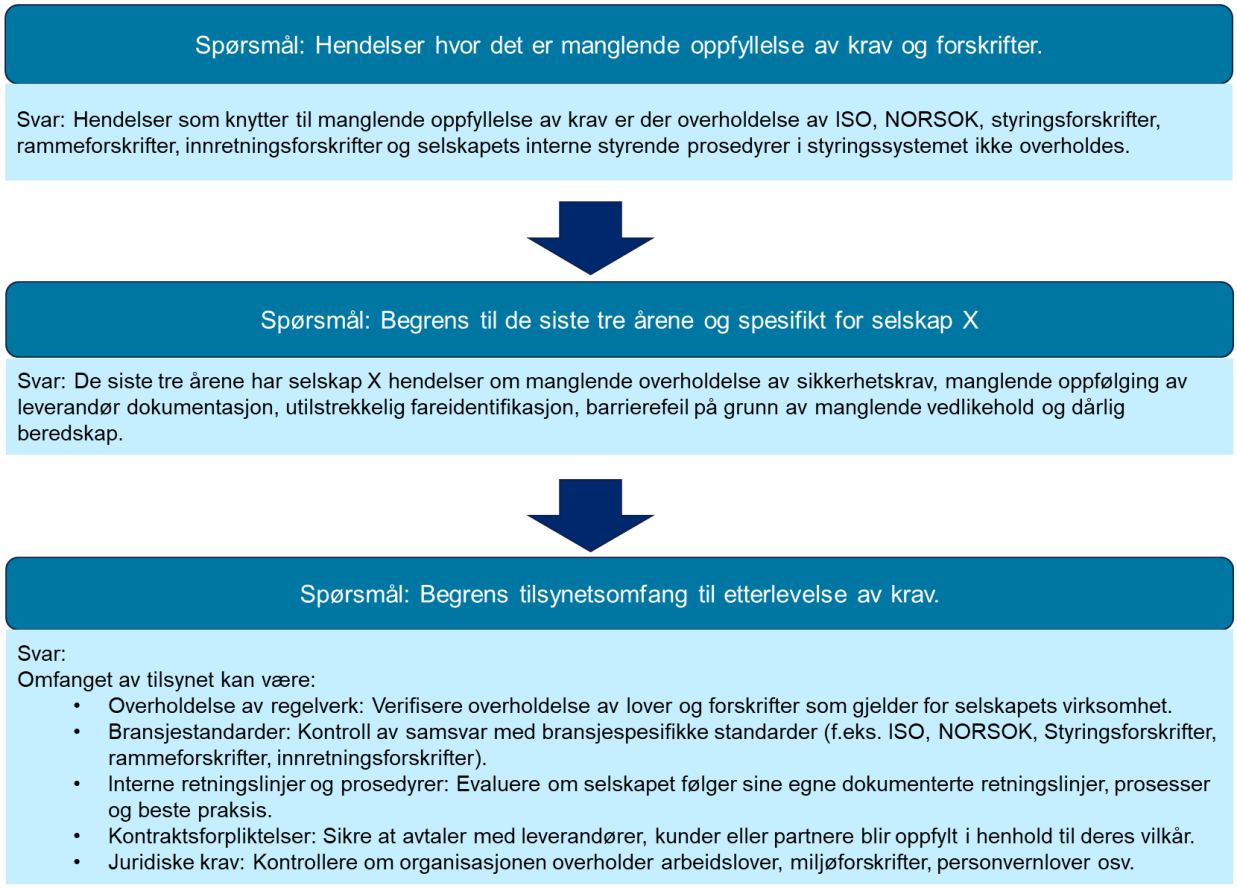
### 4.1.3 Alvorlige hendelser / høy (potensiell) konsekvens

Dette scenarioet var rettet mot spørsmål som kunne være relevant for en gransker å stille i forbindelse med en igangsatt gransking. Følgende spørsmål, samt eksempel på hvordan et meningsfullt svar kunne struktureres, ble satt opp for testing mot modellen.

1. Spørsmål: Hvilke hendelser med storulykkespotensial relatert til korrosjon har man hatt i AkerBP?
    a. Hypotetisk eksempel på svar: AkerBP har hatt 3 hendelser relatert til korrosjon. Disse var: 1) fallende gjenstand på Edvard Grieg i 2018, 2) brann/eksplosjon på Ivar Aasen i 2020, og 3) akutt utslipp på Valhall i 2020.
2. Spørsmål: Hvilke hendelser med storulykkespotensiale har Statfjord hatt de siste 10 årene relatert til fallende objekter, og hva var årsakene til disse?
    a. Hypotetisk eksempel på svar: Statfjord har hatt 2 hendelser med storulykkespotensial de siste 10 årene relatert til fallende objekter. Disse var: 1) Stillashendelse på Statfjord A i 2018. Årsakene til hendelsen var xxx. 2) Hendelse i boreområdet i 2023. Årsakene var xxx.
3. Spørsmål: Hvor mange hendelser relatert til fallende gjenstander på grunn av korrosjon har det vært i industrien de siste 10 årene, og hva var årsakene?
    a. Hypotetisk eksempel på svar: Det har vært 20 hendelser på grunn av korrosjon de siste 10 årene. Årsakene til disse var xxx.

### 4.1.4 Planlegging av tilsyn

Ved å planlegge et tilsyn er det vanlig at man ser på hendelser som har blitt meldt inn, og spesifiser f.eks. hvilket område/fag/selskap man vil se i detalj på. Under er et scenario hvor fokuset er å se på etterlevelse av krav og forskrifter hvor man stiller spørsmål eller formulerer en setning, og dermed får man et hypotetisk svar. Disse svarene kan bistå tilsynsleder med å planlegge for et tilsyn som skal utføres hos et bestemt selskap.




*Figur 3: Eksempel på scenariospørsmål og forventede svar*

### 4.1.5 Utfordringer med opprettelsen av testscenarier

En av de tidlige utfordringene var utviklingen av testscenariene. DNV satt sammen med Havtil for en online workshop hvor vi diskuterte både arbeidsflyten til Havtil-ansatte relatert til undersøkelser og planlegging av tilsyn, samt spørsmålene som hver av disse rollene muligens kunne stille LLM. Det tok litt tid å forstå hvordan man skulle formulere spørsmålene slik at LLM kunne bryte ned større komplekse spørsmål til håndterbare delspørsmål og gi det beste resultatet.

#### 4.1.5.1 Utilstrekkelig dokumentasjon

Det ble fastslått at det var vanskelig å besvare noen av de ovennevnte scenariene basert på dataene i hendelsesdatabasen. For eksempel, i scenariet for planlegging av tilsyn, hvis brukeren ønsker å spørre modellen hvilke forskrifter en bestemt operatør har brutt de siste årene, er ikke disse dataene tilgjengelige i HD. Denne typen spørsmål kunne besvares hvis en forbedret database ble opprettet med denne informasjonen hentet fra hendelsesrapporter, da denne informasjonen er tilgjengelig der. Den foreslåtte løsningen er derfor å forbedre dataene fra HD med kontekstualiserte data fra PDF-rapporter knyttet til disse sakene.

Videre er det en utfordring på grunn av utilstrekkelig dokumentasjon angående den faktiske konsekvensen. For eksempel kan datapunkter beskrive en ulykke som skjer her som å være i stillaset eller brønnen, men nesten 50 % av tilfellene er beskrevet som 'annet område', noe som i stor grad reduserer oppløsningen av dataene. I tillegg kan noen hendelser være kategorisert som DFU14 Arbeidsulykker, noe som også påvirker oppløsningen av dataene.

### 4.1.6 Løsninger

For å overvinne utfordringer rundt spørsmål i testscenarier brukte KI-teamet en god del tid sammen med DNVs fageksperter for å utvikle disse scenariene og modellere måten en menneskelig ekspert ville gå frem

for å besvare disse spørsmålene. Når vi hadde en trinnvis gjennomgang av logikken som en menneskelig ekspert ville anvende for å løse spørsmålet, kunne vi lære opp modellen.

## 4.2 Utvikling av arkitektur

### 4.2.1 Utfordringer

- Å velge passende teknologi for å muliggjøre piloten innenfor budsjett
- Å velge passende LLM-leverandør med tanke på konfidensialitetshensyn
- LLM er et raskt bevegelig domene, teknologilagene på toppen av selve språkmodellen har en tendens til å bli utdaterte raskt
- Det viste seg at det grunnleggende ved dokumentforståelse, dvs. å hente ut informasjon fra PDF-filen og deretter strukturere den i en kunnskapsgraf, er utfordrende på grunn av begrensningene i den nyeste PDF-leseteknologien.

### 4.2.2 Løsninger

- Vi bruker Python som hovedprogrammeringsspråk
- Vi valgte Azure Open AI som språkmodellleverandør
- Vi bruker relasjonsdatabase så mye vi kan og tyr til grafdatabase bare når det er nødvendig
- Vi bruker lokalt distribuert database for proof of concept.
- Vi velger å utvikle interne teknologilag på toppen av språkmodellen for å tillate fleksibilitet.

## 4.3 Opprettelse av LLM-agenter

### 4.3.1 Utfordringer

- Den største utfordringen er hvordan man kan sikre at svaret gitt av LLM er nøyaktig.
- Hvordan sikre at vi har vurdert alle kjente mulige scenarier som kan gå galt.
- Det viste seg at kolonne-/feltnavnet ikke er tydelig nok for at LLM skal forstå det.

### 4.3.2 Løsninger

- Prosesslogg vises under spørsmål- og svarøkter for en transparent prosess.
- All spørringskode produsert av LLM for å hente data fra databasen vises for å la brukeren vurdere om agenten ser på informasjonen i de riktige feltene.
- LLM-forståelse av databasen forbedres ved å endre kolonne-/feltnavn samt legge til metadata (beskrivelse) om innholdet i kolonnen.
- Når mangler ved LLM blir tydelige, legges det til ekstra sikkerhetstiltak.

## 4.4 Opprettelse og fylling av kunnskapsgrafen

### 4.4.1 Utfordringer

Noen av utfordringene med fritekstdata er at de ikke har en forhåndsdefinert struktur som kan brukes til å få innsikt i betydningen av dataene. Tekstdata er ofte tvetydige, inkonsekvente og ukontekstualiserte. Å forstå hvordan informasjonen presenteres i kontekst er viktig for å skille mellom hva som skal brukes hvor og hvordan. En annen utfordring er søk i dataene, et eksempel på dette i databasen er når en fysisk ting kalles med to synonyme navn i samme kontekst. Kvaliteten på fritekst er ofte lavere enn skjema-drevet strukturert representasjon.

Noen av utfordringene med opprettelse og befolkning av kunnskapsgrafen er å håndtere datavariasjon, enhetsekstraksjon, forståelse av relasjoner og skjema-design. Utvikling av kunnskapsgrafskjema krever domeneekspertise og validering med reelle data.

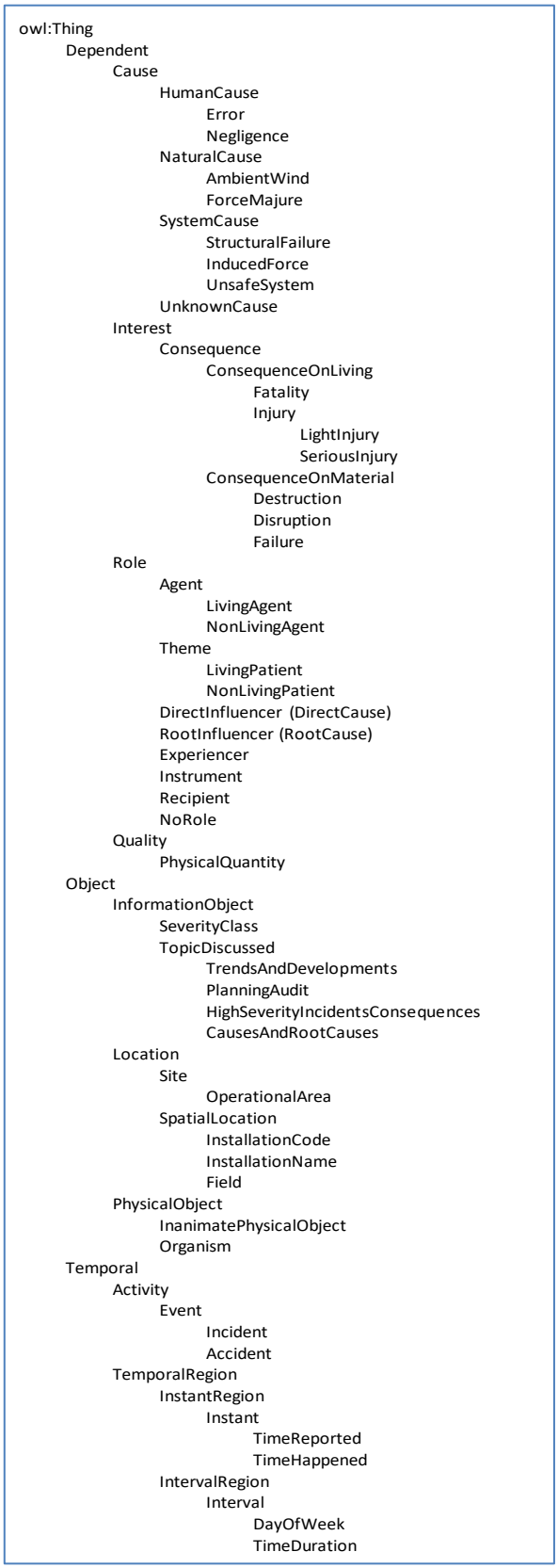
### 4.4.2 Løsninger

Den innledende løsningen på utfordringene med utvikling og befolkning av kunnskapsgrafen innebar å utvikle et kunnskapsgrafskjema i samarbeid med domeneeksperter. Skjemaet har blitt iterert over tid for å sikre at det er av tilstrekkelig kvalitet gitt tidsbegrensningene til å håndtere det meste av dataene som finnes i tekstdata og gi nok kunnskap til en LLM for å besvare bruksområdene. Denne utviklingen av kunnskapsgrafen er beskrevet i avsnitt 4.4.2.1.

Teknisk implementering av utviklingen av kunnskapsgrafskjemaet gjøres ved hjelp av lokale verktøy, mens befolkningen av faktiske data gjøres ved hjelp av LLM-assisterte metoder. Siden dataene er store, hadde vi muligheten til enten å håndtere hver rad av hendelser én om gangen eller å gi flere linjer med hendelser som ett input (kalt batcher) og kjøre ekstraksjonspipelinen. Årsaken til å bruke batcher er å redusere latensen ved ekstraksjon og redusere kostnadene, men ved validering av ekstraksjonen ble det funnet at LLM-pipelinen ikke presterte som krevd av dette prosjektet, forklart i avsnitt 4.4.2.2

#### 4.4.2.1 Utvikling av kunnskapsgraf

Den opprinnelige kunnskapsgrafen ble utviklet med hjelp av ontologer og ble validert med fageksperter. Den inneholdt 73 enheter og noen relasjoner. Den opprinnelige strukturen av enhetene vises i figur under hvor relasjonen er 'subClassOf' mellom nestede enheter.




*Figur 4: Første iterasjon av ontologien*

Den neste iterasjonen ble gjort etter at ontologer møtte fageksperter og dataforskere for gjennomgang, og det ble ansett som nødvendig å utelate noen enheter da de allerede var i strukturert format fra HD. Den resulterende ontologien inneholdt totalt 64 enheter, inkludert ontologitermene. Forbedringen som ble gjort i denne versjonen er å skille mellom en klassifisering og en navngitt enhet. Den resulterende strukturen av enhetene vises under.

- Dependent
  - Cause
    - HumanCause
      - Error
      - Negligence
    - NaturalCause
      - AmbientWind
      - ForceMajure
    - SystemCause
      - StructuralFailure
      - InducedForce
      - UnsafeSystem
    - UnknownCause
  - Description of Cause
  - Interest
    - Consequence
      - ConsequenceOnLiving
        - Fatality
        - LightInjury
        - SeriousInjury
      - ConsequenceOnMaterial
        - Destruction
        - Damage
        - Deformation
        - Disruption
        - Failure
      - NoConsequence
    - ActionsPerformedOperator
      - SecuringArea
      - Investigation
      - Inspection
      - Repair
      - Communication
      - StoppingOperation
      - NoAction
    - ActionsPerformedRegulatory
  - Role of Physical object
    - Agent
    - Patient
    - Subject
    - NoRole
  - Role of Organism
    - SAME AS ABOVE
  - Quality
    - PhysicalQuantity
      - NAMED
- Object
  - Location
    - Site
    - SpatialLocation
      - LocationFrom
      - LocationTo
  - PhysicalObject
    - InanimatePhysicalObject
      - NAMED
    - Organism
      - NAMED
- Temporal
  - Event
    - Dropped
  - Activity (Happening or During)

*Figur 5: Senere iterasjon av ontologien*

Den siste iterasjonen ble gjort etter å ha designet ekstraksjonspipelinen og eksperimentert med fritekstkolonnene for å se om enhetene fanget nok data til å gi innsikten som fageksperte trengte. Resultatene vises som datamodell i figur under, uten relasjoner.

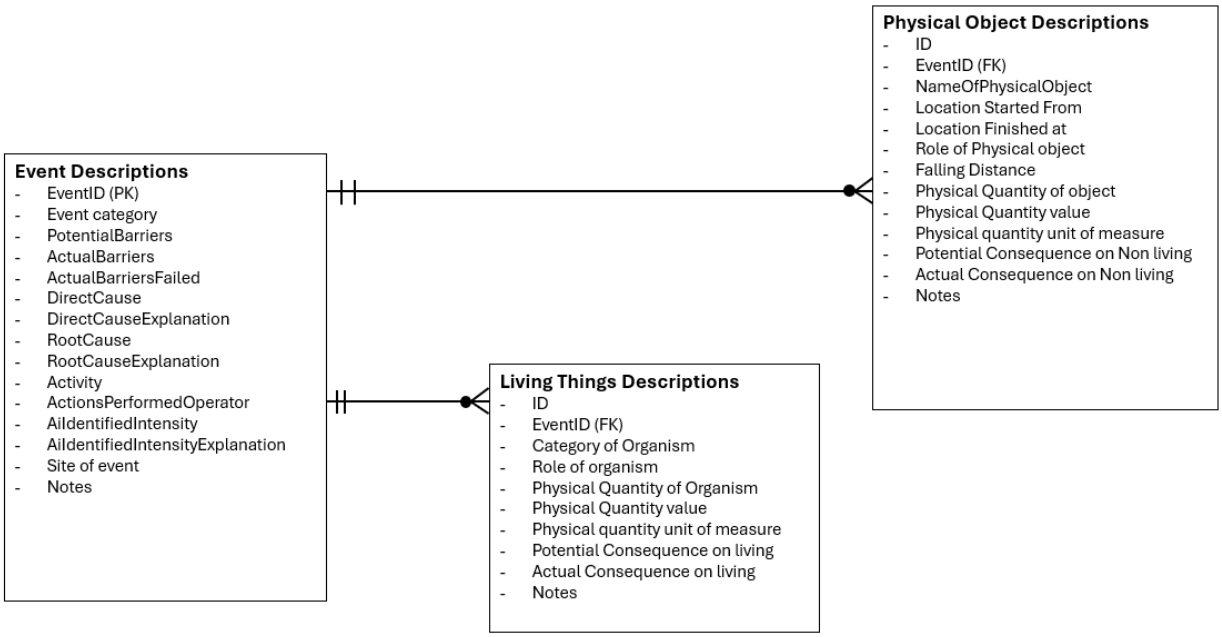



*Figure 6: Datamodellen fra den siste iterasjonen*

#### 4.4.2.2 Effektivitet av LLM i ekstraksjon av data for varierende batchstørrelser

Ulike batchstørrelser ble brukt for å forstå effektiviteten til LLM-er i utvinning av nødvendig informasjon gitt kunnskapsgrafen. Totalt 3819 hendelser ble gitt til LLM, og det resulterende antallet hendelser ved bruk av forskjellige batchstørrelser er presentert i tabellen nedenfor.

*Tabell 1: Brukte batch-størrelser og resulterende antall ekstraherte hendelser*

| Batch Size | Extracted # of events |
|---|---|
| 1 | **3819** |
| 10 | 3808 |
| 25 | 3280 |
| 40 | 2782 |

Siden antallet uttrukne hendelser reduseres når vi gir mer enn én hendelse til en LLM, er det nødvendig å bruke kun én rad når vi henter ut kunnskapsgrafpopulasjonsdata fra fritekstfelt i HD. For sammenligning og kontroll av kvaliteten på utvinningen viser vi utvinningen av fysiske objekter og deres relaterte felt i tabell 1 sammen med originalteksten for sammenligning.

*Tabell 2: Eksempel på at ekstraksjon av fysiske størrelser er lik for begge batch-størrelser.*

| Event ID | Text | Batch Size | Object | Physical Quantity | Value | Falling Distance |
|---|---|---|---|---|---|---|
| 91109 | When lifting the dump valve in column G40, the 2 T hoist and trolley fell from the monorail. The hoist (25kg) fell 6 meters and landed 1.5 meters away from person.<br>Running cat (8kg) fell about 8 meters and landed on a cable bridge nearby. Valve (970kg) fell about 1 meter and down on a pallet and was not in danger of personnel being injured her.<br>Both valve, hoist and trolley landed within sperring.<br>2 people participated in the work operation and were inside the barrier, but only one person was near the hoist. No damage to personnel, only damage to a cable in kabelgaten | 1 | Hoist | Weight | 25 | 6 meters |
| | | | Running cat | Weight | 8 | 8 meters |
| | | | Valve | Weight | 970 | 1 meter |
| | | 10 | Hoist | Weight | 25 | 6 meters |
| | | | Running cat | Weight | 8 | 8 meters |
| | | | Valve | Weight | 970 | 1 meter |

Selv om noen tilfeller ble ekskludert ved bruk av større batch-størrelser, ble ikke effektiviteten av ekstraksjonen av fysiske verdier påvirket. Det antas at årsaken til at noen tilfeller ble utelatt ved større batch-størrelser, skyldes at enkelte tilfeller har store tekstmengder sammenlignet med andre som kun inneholder en enkelt linje med tekst.

## 4.5 Strukturering og integrering

Når kolonnen "beskrivelse" er ekstrahert, delt, strukturert og kategorisert i en passende kunnskapsgraf formulert som tre ekstra relasjonstabeller. Disse tre tabellene ble deretter koblet til den opprinnelige Havtil-databasen. De strukturerte dataene lagres i en SQL-database som brukes som input til LLM-agent systemet. Resultatet av slike strukturerte data er økt nøyaktighet og utvidelse av hva som kan spørres verktøyet.

Prosessen starter med å ekstrahere "beskrivelse"-kolonnen fra databasen. Denne kolonnen inneholder detaljert informasjon som må brytes ned i mindre, mer håndterbare deler. Denne delingsprosessen hjelper til med å organisere dataene i et strukturert format. Når dataene er strukturert, kategoriseres de i forskjellige grupper, og danner en kunnskapsgraf. Denne grafen representerer forholdene og hierarkiene innen dataene, noe som gjør det lettere å forstå og analysere.

Deretter oversettes de strukturerte dataene til tre ekstra relasjonstabeller. Disse tabellene integreres deretter med den opprinnelige HAVTIL-databasen. Ved å koble disse nye tabellene med de eksisterende dataene, skaper vi et mer omfattende datasett som støtter komplekse spørringer og relasjoner. Denne integrasjonen forbedrer databasens evne til å gi detaljerte og sammenhengende informasjon.

De kombinerte dataene, nå mer robuste og strukturerte, lagres i en SQL-database. SQL-databaser er spesielt effektive for å håndtere strukturerte data fordi de støtter komplekse spørringer og sikrer dataintegritet. Denne organiserte lagringen tillater effektiv datahenting og manipulering.

SQL-databasen fungerer som input for et LLM-agentsystem. De strukturerte dataene gjør det mulig for LLM å forstå og generere svar med større nøyaktighet. LLM kan behandle den detaljerte og kategoriserte informasjonen, noe som fører til mer presise og relevante svar. Denne forbedringen utvider betydelig omfanget av spørsmål systemet kan håndtere, og forbedrer dets generelle ytelse og brukervennlighet.

## 4.6 Feil eller utfordringer med data

Feil eller utfordringer med data brukt i dette prosjektet er generelt beskrevet under spesifikke kapitler knyttet til dem. En kort oppsummering er også gitt her. Generelt er datakvaliteten i Hendelsesdatabasen begrenset, da informasjonen som sendes inn fra operatørene er ufullstendig. Ofte sender de inn kun det minimum som kreves, og på grunn av tidsbegrensninger sendes informasjonen inn ganske raskt etter hendelsen, ofte før hendelsen er gransket ferdig. Datakvaliteten påvirkes også av mangel på strenge retningslinjer for dataregistrering i Hendelsesdatabasen. Ulike brukere vektlegger visse felt, med noen felt som blir stående ufullstendige, og til tider blir informasjon skrevet inn i feil felt eller med en annen tolkning av hvilken data som skal være i hvert felt. Dette betyr at kolonner er navngitt annerledes enn innholdet. Ordet "årsak" i Hendelsesdatabasen refererer for eksempel ofte til *årsaken til sen rapportering* snarere enn årsaken til hendelsen.

## 4.7 Ekstraksjon av tekst fra forskjellige typer PDF-er

### 4.7.1 PDF-typer

PDF-filer opprettes ikke på samme måte. Det finnes minst tre typer:

- PDF-fil opprettet fra bilder eller skannet fra papirkopi.
- PDF-fil opprettet ved bruk av "skriv ut til PDF"-funksjonalitet.
- PDF-fil opprettet riktig ved bruk av "lagre som PDF"-funksjonalitet.

PDF-er opprettet fra bilder eller skannet fra papirkopi har vanligvis minimal metadata, som opprettelsesdato og skanner brukt. De inkluderer generelt ikke bokmerker eller merknader med mindre de legges til manuelt senere ved bruk av PDF-redigeringsprogramvare. Lenker er ikke funksjonelle fordi innholdet i hovedsak er et bilde. Opprinnelig er disse PDF-ene ikke søkbare, men optisk tegngjenkjenning (OCR) kan brukes for å gjøre teksten søkbar.

PDF-er opprettet ved bruk av "skriv ut til PDF"-funksjonaliteten inkluderer grunnleggende metadata som dokumenttittel, forfatter og opprettelsesdato, ofte hentet fra kildeapplikasjonen. Bokmerker og merknader opprettes ikke automatisk og må legges til manuelt hvis nødvendig. Lenker bevares vanligvis ikke, så hyperkoblinger i det opprinnelige dokumentet kan være ikke-funksjonelle i PDF-en. Imidlertid er disse PDF-ene fullt søkbare ettersom teksten bevares i konverteringsprosessen.

I kontrast har PDF-er opprettet ved bruk av "lagre som PDF"-funksjonaliteten ofte omfattende metadata, inkludert detaljerte dokumentegenskaper som tittel, forfatter, emne og nøkkelord. Hvis kildedokumentet inneholder overskrifter eller en innholdsfortegnelse, kan disse konverteres til bokmerker i PDF-en. Merknader fra kildedokumentet bevares, og hyperkoblinger og kryssreferanser er vanligvis funksjonelle. Disse PDF-ene er fullt søkbare, med teksten nøyaktig bevart.

### 4.7.2 Tilgjengelighet og funksjonalitet av overskrifter, tabeller, figurer og ligninger

PDF-er opprettet fra bilder eller skannet fra papirkopi behandler overskrifter, tabeller, figurer og ligninger som en del av bildet. Disse elementene gjenkjennes ikke som separate tekst- eller objekt, noe som gjør dem vanskelige å få tilgang til eller redigere. Overskrifter og tabeller identifiseres ikke som distinkte elementer, og figurer og ligninger er innebygd som bilder. For å gjøre disse elementene tilgjengelige, må optisk tegngjenkjenning (OCR) brukes for å konvertere bildeinnholdet til søkbar og valgfri tekst.

PDF-er opprettet ved bruk av "skriv ut til PDF"-funksjonaliteten bevarer overskrifter, tabeller, figurer og ligninger visuelt slik de vises i det opprinnelige dokumentet. Imidlertid er disse elementene ikke merket for tilgjengelighet, noe som betyr at overskrifter ikke er lett navigerbare, tabeller er statiske uten interaktive elementer, og figurer og ligninger mangler beskrivende metadata eller merker. Selv om disse PDF-ene er fullt søkbare, begrenser mangelen på strukturelle merker navigasjon og interaksjon med disse elementene.

I kontrast tilbyr PDF-er opprettet ved bruk av "lagre som PDF"-funksjonaliteten det høyeste nivået av tilgjengelighet og funksjonalitet. Overskrifter bevares og merkes, noe som gjør dem lett navigerbare og tilgjengelige. Tabeller er riktig merket, noe som tillater interaksjon med individuelle celler og rader, og figurer inkluderes med beskrivende metadata og merker. Ligninger bevares som redigerbar tekst eller matematiske uttrykk og er merket for tilgjengelighet. Disse PDF-ene er fullt søkbare og tilgjengelige, med riktig merking for alle elementer, noe som gjør navigasjon og interaksjon enkel.

I tillegg er tabeller i PDF-dokumenter vanskelige å håndtere på grunn av manglende iboende struktur, slik som metadata som nevner antall rader, kolonner, sammenslåtte celler osv. Den nåværende toppmoderne teknologien innen tabelluttrekking fungerer på vanlige tabeller, men når tabellene er mer komplekse, blir det stadig vanskeligere å trekke ut riktig struktur fra tabellene. En mulig utforskning for dette ville være å kombinere teknologiske fremskritt innen layout- og tabellstrukturuttrekking med visuelle metoder og mulig integrering av LLM-agenter for å overvåke uttrekksprosessene.

### 4.7.3 Utfordringer

Å trekke ut og dele opp informasjon fra PDF-er basert på seksjoner, underseksjoner, tabeller og andre elementer kan være utfordrende, spesielt avhengig av typen PDF.

For PDF-er opprettet fra bilder eller skannet fra papirkopi, er den primære utfordringen å konvertere bildeinnholdet til tekst ved bruk av OCR. OCR-nøyaktighet kan variere, spesielt med dårlig kvalitet på skanninger eller komplekse oppsett. Å identifisere seksjoner, underseksjoner og andre strukturelle elementer er vanskelig fordi dokumentet mangler iboende struktur. Manuell merking eller avanserte OCR-verktøy er nødvendig for å gjenkjenne disse elementene. Å trekke ut tabeller nøyaktig er utfordrende ettersom de er en del av bildet, og OCR-verktøy kan ha problemer med å skille mellom tekst og tabellelementer. Figurer og ligninger behandles som bilder, noe som gjør det vanskelig å trekke dem ut og tolke dem riktig uten manuell inngripen.

For PDF-er opprettet ved bruk av "skriv ut til PDF"-funksjonaliteten, mangler disse PDF-ene ofte strukturelle merker, noe som gjør det vanskelig å identifisere og dele opp seksjoner og underseksjoner automatisk. Selv om tabeller og figurer bevares visuelt, er de ikke merket for enkel uttrekking, noe som krever spesialiserte verktøy eller manuelt arbeid for å trekke ut disse elementene nøyaktig. Lenker og kryssreferanser bevares vanligvis ikke, noe som kan komplisere navigasjon og oppdeling basert på lenkede seksjoner. Merknader er ikke inkludert med mindre de legges til manuelt, noe som kan begrense konteksten tilgjengelig for oppdeling.

I kontrast bevarer PDF-er opprettet ved bruk av "lagre som PDF"-funksjonaliteten struktur og merker, men svært komplekse oppsett kan fortsatt utgjøre utfordringer for automatisk oppdeling. Kvaliteten på PDF-en avhenger av programvaren som brukes til å opprette den, og noen applikasjoner støtter kanskje ikke alle funksjoner fullt ut, noe som kan føre til potensielt tap av strukturell informasjon. Større filstørrelser på grunn av detaljerte metadata og merking kan påvirke ytelsen til uttrekksverktøy, noe som gjør prosessen tregere. I tillegg kan noen PDF-er ha begrensninger på redigering eller uttrekking av innhold, noe som kan hindre oppdelingsprosessen.

For å oppnå dette målet kan avanserte PDF-uttrekksverktøy som støtter OCR, strukturell gjenkjenning og merking være nødvendige. I tillegg kan manuell inngripen være nødvendig for å sikre nøyaktighet, spesielt for komplekse dokumenter.

### 4.7.4 Innsats

Bruk av Python og biblioteker som PyMuPDF kan effektivisere prosessen med å trekke ut og dele opp informasjon fra forskjellige typer PDF-er. For PDF-er opprettet fra bilder eller skannet fra papirkopi, er den primære utfordringen å konvertere bildeinnholdet til tekst ved bruk av OCR. PyMuPDF kan kombineres med OCR-biblioteker som Tesseract for å håndtere dette, men prosessen kan være tidkrevende og kan kreve manuell korreksjon for nøyaktighet. Betydelig manuelt arbeid er nødvendig for å merke og strukturere dokumentet, inkludert identifisering av overskrifter, tabeller, figurer og ligninger. Python-skript kan hjelpe med å automatisere noe av dette, men manuell inngripen er ofte nødvendig. Å trekke ut tabeller, figurer og ligninger nøyaktig er utfordrende og krever ofte spesialiserte verktøy eller manuell inngripen. Biblioteker

som OpenCV kan hjelpe med å identifisere og trekke ut disse elementene. Totalt sett innebærer denne metoden et høyt nivå av innsats på grunn av behovet for omfattende manuell behandling og korreksjon, selv med hjelp av Python-biblioteker.

For PDF-er opprettet ved bruk av "skriv ut til PDF"-funksjonaliteten, mangler disse PDF-ene ofte strukturelle merker, så identifisering og oppdeling av seksjoner og underseksjoner krever manuelt arbeid eller avanserte verktøy. PyMuPDF kan trekke ut tekst, men ytterligere behandling er nødvendig for å identifisere strukturen. Å trekke ut tabeller og figurer er mindre utfordrende enn med skannede PDF-er, men krever fortsatt noe manuelt arbeid eller spesialisert programvare. Biblioteker som Camelot eller Tabula kan hjelpe med tabelluttrekking. Lenker og merknader bevares ikke, noe som krever ekstra manuelt arbeid for å gjenopprette disse elementene hvis nødvendig. PyMuPDF kan hjelpe med å identifisere og gjenskape lenker. Denne metoden krever et moderat nivå av innsats, med noe manuelt arbeid nødvendig for å strukturere og trekke ut informasjon nøyaktig, selv med Python-biblioteker.

I kontrast bevarer PDF-er opprettet ved bruk av "lagre som PDF"-funksjonaliteten struktur og merker, noe som gjør det lettere å identifisere og dele opp seksjoner, underseksjoner, tabeller, figurer og ligninger. PyMuPDF kan effektivt trekke ut denne strukturerte informasjonen. Mange avanserte PDF-verktøy og biblioteker kan håndtere disse PDF-ene effektivt, noe som reduserer behovet for manuell inngripen. PyMuPDF, sammen med andre biblioteker som PDFMiner, kan automatisere mye av uttrekksprosessen. Håndtering av svært komplekse oppsett kan fortsatt kreve noe manuelt arbeid, men generelt er prosessen mer strømlinjeformet med Python-skript. Denne metoden innebærer minst innsats på grunn av bevaring av struktur og merker, noe som gjør det lettere å automatisere uttrekks- og oppdelingsprosessen ved bruk av Python-biblioteker.

Oppsummert krever uttrekking og oppdeling av informasjon fra PDF-er opprettet fra bilder eller skannede dokumenter den høyeste innsatsen, mens PDF-er opprettet ved bruk av "lagre som PDF"-funksjonaliteten krever minst innsats. PDF-er opprettet ved bruk av "skriv ut til PDF"-funksjonaliteten faller et sted imellom. Bruk av Python og biblioteker som PyMuPDF, Tesseract, Camelot og andre kan betydelig hjelpe med å automatisere og strømlinjeforme prosessen, selv om noe manuell inngripen fortsatt kan være nødvendig for å sikre nøyaktighet og fullstendighet.

#### 4.7.4.1 Teknisk

Som i andre dokumenter kan en PDF bestå av tekst, figurer/bilder, tabeller og ligninger. Noen ganger viser et bilde en tabell, som kan være skannet. Ideelt sett bør data-/informasjonsekstraksjon fra PDF-filer innebære:

1. Ekstraksjon av all tekst og riktig kategorisering som overskrift eller underoverskrift, tekstinnhold, tabelltekst, tabelloverskrift, tabellinnhold, figurtekst og ligning.
2. Ekstraksjon av alle bilder og lagring av dem med metadata (som bildetekst, seksjon/underseksjon og sidetall).
3. Ekstraksjon av alle ligninger og lagring av dem sammen med metadata (som formål, ligningsnummer, seksjon/underseksjon og sidetall).
4. Konvertering av ligninger til LaTeX-format.

#### 4.7.4.2 Konfidensialitet

Hvis rapporten er konfidensiell, må det utvises forsiktighet før den sendes til LLM API. Man kan utføre anonymisering. Dette er imidlertid ikke enkelt og kan kreve manuell inngripen. Fullt automatisert anonymisering kan kreve LLM, noe som igjen betyr å sende den ut til LLM API. Derfor er den mest realistiske løsningen å kjøre LLM lokalt. Dette bør undersøkes nærmere i evt. videre arbeid.

## 4.8 Risikoer

Det er en rekke utfordringer og risikoer med KI-modeller som aktører i petroleumsvirksomheten må være oppmerksomme på og ta høyde for i sine risikovurderinger og risikoreduserende tiltak.

1. **Datasikkerhet og personvern**: KI-modeller krever store mengder data for å trenes effektivt. Dette kan inkludere sensitive og konfidensielle opplysninger. Risikoen for datalekkasjer og uautorisert tilgang til data er betydelig, og det er viktig å implementere robuste sikkerhetstiltak for å beskytte dataene.
2. **Bias og diskriminering**: KI-modeller kan arve skjevheter fra treningsdataene, noe som kan føre til diskriminerende beslutninger. Dette er spesielt kritisk i sikkerhetskritiske industrier som petroleumsvirksomheten, hvor feilaktige beslutninger kan ha alvorlige konsekvenser. Det er viktig å sikre at treningsdataene er representative og fri for skjevheter.
3. **Forklarbarhet og transparens**: Mange KI-modeller, spesielt dype nevrale nettverk, fungerer som "svarte bokser" hvor det er vanskelig å forstå hvordan de kommer frem til sine beslutninger. Dette kan være problematisk i en industri hvor det er krav til dokumentasjon og forklaring av beslutningsprosesser. Det er viktig å utvikle metoder for å gjøre modellene mer forklarbare og transparente. I vårt verktøy har vi inkludert en prosesslogg for å kommunisere prosessen LLM-modellen har brukt for å komme frem til svaret.
4. **Avhengighet av teknologi**: Økt bruk av KI kan føre til en overavhengighet av teknologi, hvor menneskelig ekspertise og vurdering blir nedprioritert. Dette kan være risikabelt i situasjoner hvor teknologien svikter eller møter uforutsette problemer. Det er viktig å opprettholde en balanse mellom teknologibruk og menneskelig innsikt.
5. **Regulatoriske og juridiske utfordringer**: Bruken av KI i petroleumsvirksomheten kan reise nye regulatoriske og juridiske spørsmål, spesielt knyttet til ansvar og overholdelse av sikkerhetsstandarder. Det er viktig å holde seg oppdatert på gjeldende lover og forskrifter, og sikre at KI-løsningene er i samsvar med disse.

Ved å være oppmerksom på disse risikoene og implementere passende risikoreduserende tiltak, kan aktører i petroleumsvirksomheten dra nytte av KI-teknologiens fordeler samtidig som de minimerer potensielle ulemper.

### 4.8.1 Sårbarheter for bevisst manipulering

#### 4.8.1.1 Bevisst manipulering relatert til sikkerheten til databasen

En stor bekymring er SQL-injeksjonsangrep, hvor brukere kan forsøke å injisere skadelig SQL-kode gjennom input for å manipulere databasen. For å redusere denne risikoen bør man implementere inputvalidering og sanitering, og bruke parameteriserte spørringer eller forhåndsdefinerte utsagn for å forhindre SQL-injeksjon. Et annet problem er privilegie-eskalering, hvor brukere kan forsøke å oppnå høyere tilgangsnivå til databasen enn tiltenkt. Å sikre at LLM opererer med minst mulig nødvendige privilegier og bruke rollebasert tilgangskontroll (RBAC) kan bidra til å begrense hvilke handlinger LLM kan utføre.

Kommandoinjeksjon er en annen sårbarhet, hvor brukere kan forsøke å utføre vilkårlige kommandoer ved å manipulere input. For å forhindre dette bør man validere og sanere all input, og bruke en hviteliste-tilnærming hvor kun spesifikke, sikre kommandoer er tillatt. Dataeksfiltrering er også en risiko, hvor brukere kan forsøke å trekke ut sensitiv data fra databasen. Implementering av strenge dataadgangspolicyer og overvåking og logging av databasespørringer kan hjelpe med å oppdage uvanlige tilgangsmønstre.

Denial of Service (DoS)-angrep, hvor brukere kan forsøke å overbelaste systemet med overdrevne spørringer, slik at det blir tregt eller krasjer, kan reduseres ved å begrense antall spørringer en bruker kan gjøre i en gitt tidsperiode og implementere grenser for spørringskompleksitet. Semantisk manipulering, hvor brukere kan lage input som subtilt endrer den tiltenkte spørringslogikken, kan adresseres ved å bruke teknikker for naturlig språkforståelse for bedre å forstå brukerens hensikt og implementere robust testing for å identifisere og håndtere kanttilfeller.

Uautoriserte skjemaendringer, som å legge til eller fjerne kolonner, kan forhindres ved å begrense kommandoer for skjemaendring og implementere en gjennomgangs- og godkjenningsprosess for alle skjemaendringer. Riktig logging og overvåking er essensielt for å spore alle interaksjoner med databasen og bruke anomali-deteksjon for å identifisere mistenkelige aktiviteter. I tillegg kan sterke brukerautentiserings-

og autorisasjonsmekanismer, som flerfaktorautentisering, forhindre uautoriserte brukere fra å få tilgang til systemet.

Til slutt kan regelmessige sikkerhetsrevisjoner og penetrasjonstesting hjelpe med å identifisere og fikse sårbarheter, og sikre at LLM-chatboten forblir sikker mot bevisst manipulering. Ved å adressere disse potensielle sårbarhetene kan sikkerheten til LLM-chatboten forbedres betydelig, og beskytte databasen mot bevisst manipulering.

#### 4.8.1.2 Bevisst manipulering relatert til manipulerte chatbot-svar

En annen viktig bekymring er brukere som forsøker å manipulere chatbotens svar for å reflektere det de ønsker, i stedet for hva databasen faktisk inneholder. Dette kan undergrave integriteten og påliteligheten til informasjonen som gis.

For å adressere dette problemet bør strenge regler for spørringsvalidering implementeres for å sikre at spørringene generert av LLM nøyaktig reflekterer brukerens hensikt og er konsistente med databaseskjemaet. Å forbedre LLMs evne til å forstå konteksten og hensikten bak brukerspørringer gjennom avanserte teknikker for naturlig språkbehandling kan også hjelpe.

I tillegg kan et verifikasjonslag som kryssjekker resultatene returnert av databasen mot forventede utfall identifisere avvik og sikre nøyaktige svar. Innføring av en tilbakemeldingsmekanisme hvor brukere kan rapportere avvik eller problemer med svarene kan ytterligere forbedre systemet og adressere manipuleringsforsøk. Å opprettholde åpenhet ved å logge alle spørringer og svar tillater revisjon og gjennomgang av interaksjoner for å oppdage eventuelle manipuleringsmønstre, og anomali-deteksjonsalgoritmer kan identifisere uvanlige spørringsmønstre.

Å begrense brukernes evne til å påvirke chatbotens svar gjennom strenge tilgangskontroller sikrer at kun autoriserte brukere kan gjøre endringer i databasen eller chatbotens konfigurasjon. Regelmessige oppdateringer og revisjoner av LLM og databasen er også essensielt for å sikre at de er sikre og oppdaterte, og hjelper med å identifisere og adressere eventuelle sårbarheter som kan utnyttes for manipulering. Ved å implementere disse strategiene kan integriteten til chatbotens svar opprettholdes, og sikre at informasjonen som gis er nøyaktig og reflekterer det faktiske innholdet i databasen.

## 4.9 Tidlige spørsmål og nøyaktighetssjekker

Modellen ble testet sammen med fagpersoner flere ganger gjennom prosjektet. Noen av disse testene er inkludert i vedlegget for å fremheve noen av de utførte testene, utfordringene som ble møtt, og lærdommene som ble tatt med for å forbedre modellen. Se appendix for eksempler.

- Modellen var svært god på statistiske operasjoner og ga ekstremt nøyaktige svar når det var klare tilsvarende data i databasen.
- Modellen viste at den kunne forstå synonymer for det meste, men det var visse situasjoner hvor den virket kresen, noe som ble et punkt for videre undersøkelse.
- Variasjoner i stavemåte ble også notert. Modellen hadde ingen problemer med vanlige ord som var feilstavet, men når det gjaldt feilstavede navn, var det ingen toleranse.
- Variasjon i stavemåten av firmanavn var fleksibel; for eksempel ble flere måter å skrive ConocoPhillips på inkludert i et søk, men et feilstavet feltnavn førte til at modellen ikke returnerte noe resultat.
- Modellen kunne håndtere både engelsk og norsk input uten problemer. Søk kunne gjøres på begge språk, og resultatet ville matche inputspråket med mindre annet var spesifisert.

### 4.9.1 Uforklarlige resultater og gjennomsiktige prosesser

Den LLM-baserte chatboten består av en rekke LLM-agenter og LLM-sikkerhetsbarrierer som jobber sammen for å gi svar på ethvert bruker-spørsmål. Resultatene eller svarene kan noen ganger være

uforklarlige, og prosessene bak kulissene kan være ugjennomsiktige. For å gjøre prosessen mer transparent og forklarlig, er følgende tiltak iverksatt:

1. Implementere omfattende logging av alle interaksjoner, inkludert de genererte SQL-spørringene, svarene fra databasen og beslutningene tatt av hver LLM-agent. Dette gir en detaljert revisjonsspor som kan gjennomgås for å forstå hvordan et bestemt svar ble avledet.
2. Bruke forklarbare KI-teknikker for å gi innsikt i hvordan LLM-agentene kommer frem til sine beslutninger. Dette kan inkludere å generere forklaringer på hvorfor visse spørringer ble konstruert på en bestemt måte eller hvorfor spesifikke data ble hentet.

## 4.10 Menneske-maskin-interaksjon

Menneske-maskin-interaksjon («HMI») omfatter både hvordan verktøyet presenterer informasjon til brukeren og hvordan brukeren gir input til verktøyet. Verktøyet har innebygde sikkerhetsmekanismer for først å evaluere om spørsmålet er gyldig. Modellen kan også foreslå noen spørsmål tilbake til brukeren.

En funksjon vi ønsker å utforske er å trene modellen svært godt på spesifikke typer spørsmål slik at modellen kan veilede brukeren til å stille passende spørsmål. Prototypen som ble bygget inkluderte en minnekomponent, slik at modellen husker det forrige spørsmålet og påfølgende spørsmål kan grave dypere inn i samme tema. Dette kan imidlertid være uheldig til tider hvis brukeren ikke husker å tilbakestille chatten. For eksempel, hvis du spør om antall hendelser som oppstår på et bestemt felt og deretter spør om antall hendelser som involverer et bestemt element, for eksempel "bolter", vil ditt andre svar bli filtrert av resultatet av det første, noe som kanskje ikke alltid er brukerens hensikt.

For å forbedre brukeropplevelsen, er det viktig å utvikle klare retningslinjer og opplæring for brukerne om hvordan de best kan bruke verktøyet. Dette inkluderer å forstå hvordan man tilbakestiller konteksten og hvordan man formulerer spørsmål for å få de mest nøyaktige svarene. Videre kan det være nyttig å implementere visuelle eller tekstuelle indikatorer som minner brukeren om den nåværende konteksten og gir muligheter for å tilbakestille eller endre søkekriteriene.

Gjennom kontinuerlig evaluering og tilbakemelding fra brukerne, kan vi justere og forbedre interaksjonsmodellen for å sikre at den oppfyller brukernes behov og forventninger på en effektiv og brukervennlig måte.

For å sikre at verktøyet fungerer innenfor trygge rammer, har vi implementert flere sikkerhetsmekanismer, eller "guardrails". Disse mekanismene hjelper til med å veilede verktøyet og brukeren, og sørger for at interaksjonene forblir relevante og trygge. For eksempel, verktøyet evaluerer kontinuerlig gyldigheten av spørsmålene som stilles, og kan foreslå forbedringer eller alternative spørsmål. Dette bidrar til å forhindre misforståelser og sikrer at brukeren får mest mulig nøyaktige og nyttige svar. I tillegg kan verktøyet gi advarsler eller veiledning når det oppdager potensielt problematiske eller sensitive spørsmål, noe som ytterligere styrker sikkerheten og påliteligheten i bruken.

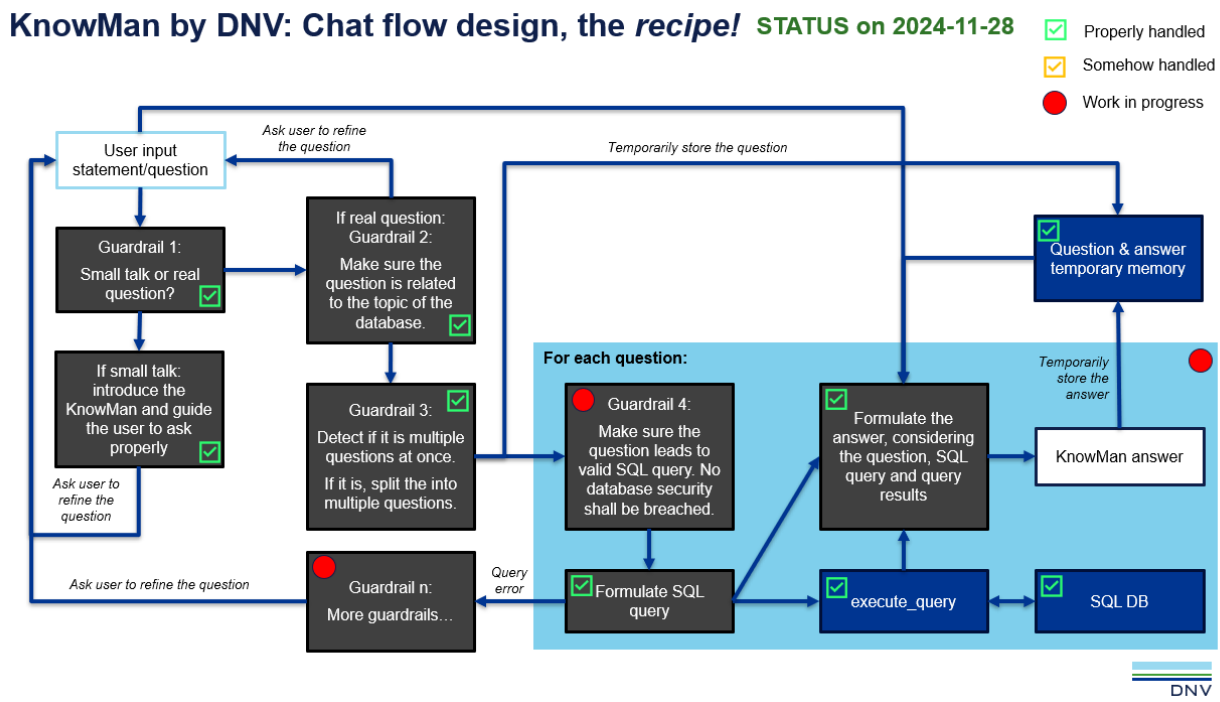



*Figure 7: Struktur på sikkerhetsmekanismer, eller "guardrails"*

# 5 ANBEFALING FOR VIDEREFØRING AV DENNE STUDIEN

Dette prosjektet har møtt en rekke utfordringer, som beskrevet i denne rapporten. Mange av disse utfordringene har blitt overvunnet, men det gjenstår fortsatt noen utfordringer som DNV ikke har vært i stand til å implementere på grunn av tidsbegrensninger. Noen av disse ideene for å adressere de gjenværende utfordringene er beskrevet i de følgende seksjonene. Til tross for de eksisterende begrensningene, er det viktig å anerkjenne de betydelige fremskrittene som allerede er gjort, og å fortsette å arbeide mot å finne løsninger på de gjenværende problemene.

## 5.1 Automatisk vurdering av LLM-resultater

En av de største utfordringene ved evaluering av maskinlæring generelt og LLM spesielt, er behovet for omfattende menneskelige annotasjoner for å vurdere kvaliteten på de genererte svarene. Dette er absolutt ikke praktisk og begrenser bruken under verktøyets driftsfase. Derfor er en automatisert evalueringsfunksjon avgjørende. Videre studier på dette emnet er nødvendig.

Med fokus på LLM-applikasjonen beskrevet i denne rapporten, altså en chatbot drevet av LLM og en database, hvor LLM utfører forståelse av menneskelige spørsmål og oversettelse av spørsmålet til SQL-spørringskode. Et Python-program kjører deretter SQL-koden og mottar resultatene. Spørringsresultatene sendes tilbake til LLM for å formulere det endelige svaret som presenteres for brukeren. Her er noen punkter som kan tjene som retning for hva man bør vurdere under automatisert evaluering:

1. Sikre at SQL-spørringen overholder databaseskjemaet ved å sjekke tabellnavn, kolonnenavn og datatyper. Dette hjelper med å verifisere at tabellene og kolonnene som refereres i spørringen eksisterer, og at operasjonene er kompatible med datatypene.
2. Ekstrahere tabellnavn fra SQL-spørringen og nøkkelordene som brukes, berike dem med beskrivelsen av kolonnenavnene, og vurdere om informasjonen sannsynligvis gir svaret på spørsmålet. En LLM-agent kan utformes for å utføre vurderingen.
3. Nøyaktigheten av svaret kan vurderes ved metoder som ligner på RAGAS (Retrieval Augmented Generation Assessment), som er et rammeverk for referansefri evaluering.

## 5.2 Mulige integrasjoner med andre datakilder

Det har blitt bemerket at noen av spørsmålene som Havtil ønsker å stille verktøyet, ikke kan besvares av Hendelsedatabasen alene. Ved å integrere andre datakilder vil verktøyet bli langt mer nyttig, og i stand til å besvare spørsmål som går utover begrensningene i dataene innenfor HD. Dette vil ikke bare øke verktøyets funksjonalitet, men også gi brukerne en mer omfattende forståelse av de komplekse sammenhengene og mønstrene som påvirker deres arbeid. Videre vil en slik integrasjon kunne bidra til mer informerte beslutninger og bedre risikostyring, noe som er avgjørende for å oppnå høyere sikkerhetsstandarder og effektivitet i operasjonene. De følgende seksjonene vil utdype hvordan denne integrasjonen kan gjennomføres og hvilke fordeler den vil medføre.

### 5.2.1 Inkludere granskningsrapporter

Ved å inkludere granskningsrapporter som en del av modellens datakilder, vil modellen få tilgang til langt mer dyptgående data enn det som finnes i Hendelsedatabasen. Undersøkelsesrapporter inneholder ofte detaljerte beskrivelser av hendelsen, inkludert den direkte årsaken, rotårsaken, eventuelle brudd på forskrifter eller prosedyrer, og de tiltakene som ble iverksatt. Ved å supplere Hendelsedatabasen med denne rikere informasjonen, vil Havtil kunne stille noen av de mer komplekse spørsmålene som ble diskutert under den digitale workshopen.

Denne integrasjonen vil ikke bare forbedre modellens evne til å gi presise og omfattende svar, men også styrke brukernes innsikt i de underliggende årsakene til hendelser og hvordan de kan forebygges i fremtiden. Dette vil være et viktig skritt mot å forbedre sikkerheten og effektiviteten i operasjonene, og sikre at alle relevante aspekter blir vurdert i beslutningsprosessen. De følgende seksjonene vil beskrive hvordan denne integrasjonen kan gjennomføres og hvilke spesifikke fordeler den vil medføre for brukerne.

### 5.2.2 Inkludere rapporter fra operatorene

Selv om Havtil samler informasjon fra industrien, finnes det ofte langt mer dyptgående data hos operatørene. Våre innledende diskusjoner med operatørene viser deres vilje til å dele lærdommer, men beskyttelsen av informasjon fra et juridisk perspektiv er en hindring som må overvinnes. Dette verktøyet kunne bli ekstremt kraftig i å analysere trender på tvers av industrien dersom det fikk tilgang til begrensede data som konsekvenser, årsaker, rotårsaker og tiltak.

Ved å overvinne disse juridiske hindringene og sikre tilgang til denne typen data, kan vi oppnå en mer helhetlig forståelse av hendelser og deres underliggende årsaker. Dette vil ikke bare forbedre verktøyets evne til å identifisere og analysere trender, men også styrke operatørenes evne til å implementere effektive forebyggende tiltak. De følgende seksjonene vil utdype hvordan vi kan navigere de juridiske utfordringene og hvilke spesifikke fordeler en slik tilgang vil medføre for både verktøyet og industrien som helhet.

### 5.2.3 Innsyn i tilsyn

DNV har kategorisert et datasett fra alle hendelsesrapportene som er tilgjengelig på Veracity i form av et Power BI-dashbord. Dataene i dette datasettet vil være svært verdifulle for å forbedre modellens svar på mange spørsmål, spesielt rundt hvilke forskrifter som er brutt. Å teste verdien av dette datasettet for noen spørsmål vil være et interessant eksperiment.

### 5.2.4 Bredere spekter av hendelser

For videre utvikling av teknologien er det et forslag å utvide omfanget til å omfatte et bredere spekter av hendelser utover det valgte temaet, ved å utnytte innsikt og forbedringer fra dette prosjektet. Dette vil innebære å inkludere flere typer data og hendelser, noe som vil gi en mer omfattende og detaljert forståelse av de ulike faktorene som påvirker sikkerheten og effektiviteten i industrien.

Ved å bygge videre på de erfaringene og lærdommene som er oppnådd gjennom dette prosjektet, kan vi utvikle et verktøy som ikke bare er mer robust og allsidig, men også bedre tilpasset de komplekse behovene til brukerne. Dette vil kreve en systematisk tilnærming til datainnsamling og analyse, samt et tett samarbeid med operatørene for å sikre at alle relevante data blir inkludert og beskyttet på en forsvarlig måte.

## 5.3 Potensielt videre industri-samarbeid

Samtidig med gjennomføringen av dette prosjektet ble det utført en rekke dialog-møter med operatører med fokus på utfordringer og mulige nytteverdier av en slik lærings-løsning. Det er observert ganske like utfordringer og nytteverdier for operatørene, spesielt med videre berikelse av strukturerte, søkbare og analyserbare hendelsesdata koblet til deres egne systemer og databaser.

Det ble vist god interesse for deltakelse i potensielt videre arbeid. Største hindring er deling av hendelsesdata med hensyn til forhold som konfidensialitet og omdømme. Mens, det kan være lettere å dele anonymiserte læringsdata i etterkant i en læringsprosess.

Enkelte av operatørene kan tenke seg å kjøre løsningen som er utviklet i dette prosjektet på sine egne lokale data først. Det eneste som deles mellom selskapene er “søkemotoren” med LLM og kunnskapsgrafer. På denne måten kan en felles “søkemotor” og lærings-tjeneste deles og utvikles videre mellom industri-parter, med mål om forbedret læring innad i og på tvers av organisasjoner.

Et videre arbeid kan ta form som et felles industriprosjekt (JIP) med fokus på delt “søkemotor” og læringstjeneste. Samtidig som det kjøres flere en-til-en piloter lokalt med operatører som stiller til rådighet egne data, systemer og ekspertise. Det er også viktig å være bevisste på hvilke områder og data som velges for dette arbeidet, slik at det fasiliteres læring og forbedring på tvers av operatører og fagområder.

I løpet av dette prosjektet er det pekt på et mulig felles industri-initiativ. Det foreslås derfor videre arbeid med planlegging og tilrettelegging for et felles-møte med Havtil og operatører tentativt i mars 2025.

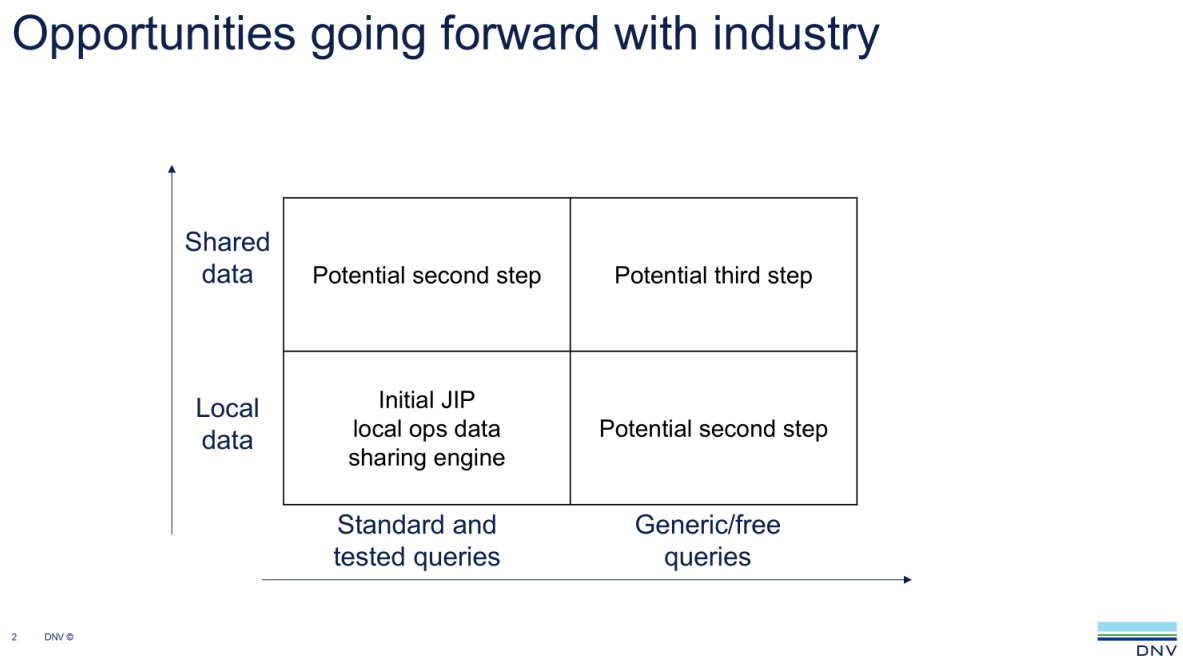
En trinnvis lærings/forbedrings-prosess og vei framover for løsningen kan sees i to hoved-dimensjoner.

- Datagrunnlag: lokale grunndata (operatør) og delte læringsdata (industri, Havtil)

- “Søkemotor”: standard testede spørsmål, i forhold til åpne og generiske spørsmål

I dette prosjektet er det laget et startpunkt for løsnings-fundament med standard testede spørsmål på data tilgjengelig gjort fra Havtil.

Dette fundamentet vil også legge grunnlag for videre veivalg sammen med industrien og Havtil.




*Figur 8: Trinnvis lærings- og forbedringsprosess i to hoveddimensjoner*

# 6 APPENDIX

## 6.1 Querying the LLM

Below are some of the early-stage questions we asked the model to look for accuracy and check the logic that the model was using to provide an answer. The results from the model were checked for accuracy by DNV's model developers and subject matter experts against the data source.

### 6.1.1 Statistics

In general statistics checks the results were checked and found to be correct.

The results were all severity 4, however the model did not try to determine the most serious any more than the highest assigned severity value.

### 6.1.2 Synonyms

Results shows that using synonyms like "dangerous" instead of "severe" also works.

### 6.1.3 Variations in spelling

The result was missing a single case from severity 2 under the operator's name "Conoco", otherwise correct.

Searching for Conoco provided the correct answer. Testing also shows common spelling errors like "teh" do not affect the model.

Correct answer for all cases containing the text "Conoco". This series of tests highlighted the need for better data quality in Hendelsedatabasen. Many operator names have changed over time or are recorded differently by those entering data.

*Tabell 3: Eksempler på flere navn for operatører*

| | | | | | |
|---|---|---|---|---|---|
| Conoco | Det norske Oil AS | E.ON Ruhrgas | ENI | Equinor | Esso |
| ConocoPhillips | DNO | E.ON Ruhrgas Norge AS | Eni Norge AS | Equinor Energy AS | EssoMobil |
| CONOCOPHILLIPS SKANDINAVIA | DNO NORGE AS | | | Statoil | Exxon Mobil Corporation |
| ConocoPhillips Skandinavia AS | DNO North Sea (Norge) AS | | | Statoil Petroleum | ExxonMobil |
| Phillips | | | | Statoil Petroleum AS | ExxonMobil Exploration and Pro |
| | | | | StatoilHydro | ExxonMobil Exploration and Production Norway AS |

Variations in names highlight the need for data within HD to be cleaned or alternatively recognition of different naming conventions to be built into the model.

### 6.1.4 Specificity

**User**: Provide the number of events for high severity category for each operator for year 2024.

**KnowMan**:

There are no events recorded for the high severity category (severity level 5) for any operator in the year 2024.

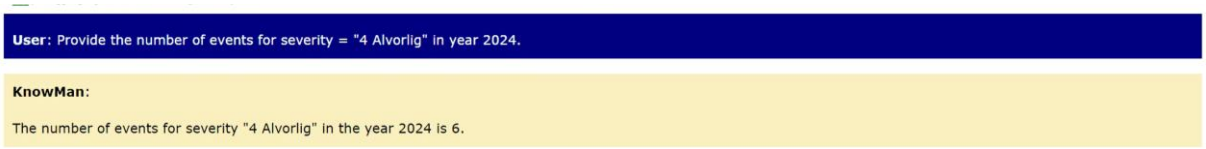
The answer is correct, however as there are very few cases with severity level 5, severity level 4 (serious) should be included. This may need to be built into the model.

**User**: Provide the number of events for severity = 4 for each operator for year 2024.

**KnowMan**:

The result indicates that there are no events with severity = 4 for any operator in the year 2024. Therefore, the answer to your question is that there are zero events for severity = 4 for each operator for the year 2024.

This is incorrect as there are cases with severity level 4 in 2024.




This answer is correct, it can be seen here that the model is sensitive to exact wording in this case. Severity = 4 was considered different than severity = 4 Alvorlig. The use of the equals sign forces the model to look for exact wording which could often not be the intention of the user.

There were more incidents than expected, however it was found that the result included “Greater Ekofisk”. The user should be aware when creating a prompt about how specific their desired result should be. For example; when searching for the company ConocoPhillips, the user may desire results to include Conoco and ConocoPhillips Scandinavia. However in the above case the user may only wish the include the results of the Ekofisk field and not Greater Ekofisk. The user could specify “field = Ekofisk” to force the model to look for the exact working.

In this incident, a typo in the name of the field affected the results. It would appear spelling errors usually don’t affect the result unless it is the spelling of a name.

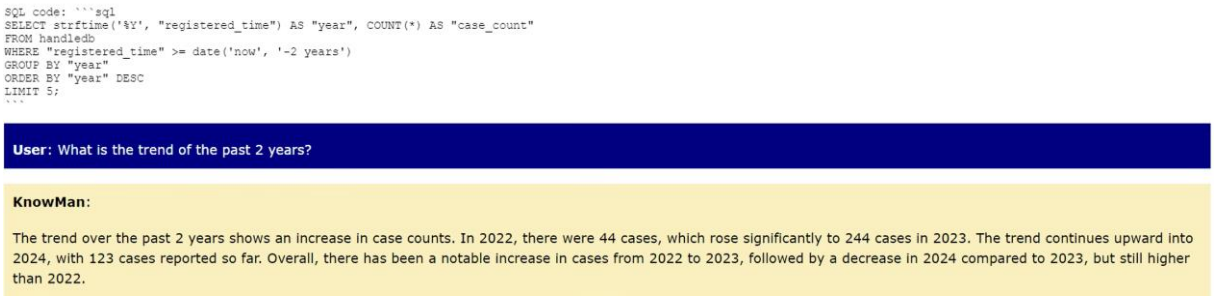
### 6.1.5 Language

Wrench is not mentioned in the description field once. But “skiftenøkkel” is mentioned often. This indicates the model is able to recognise words that are written in other languages.

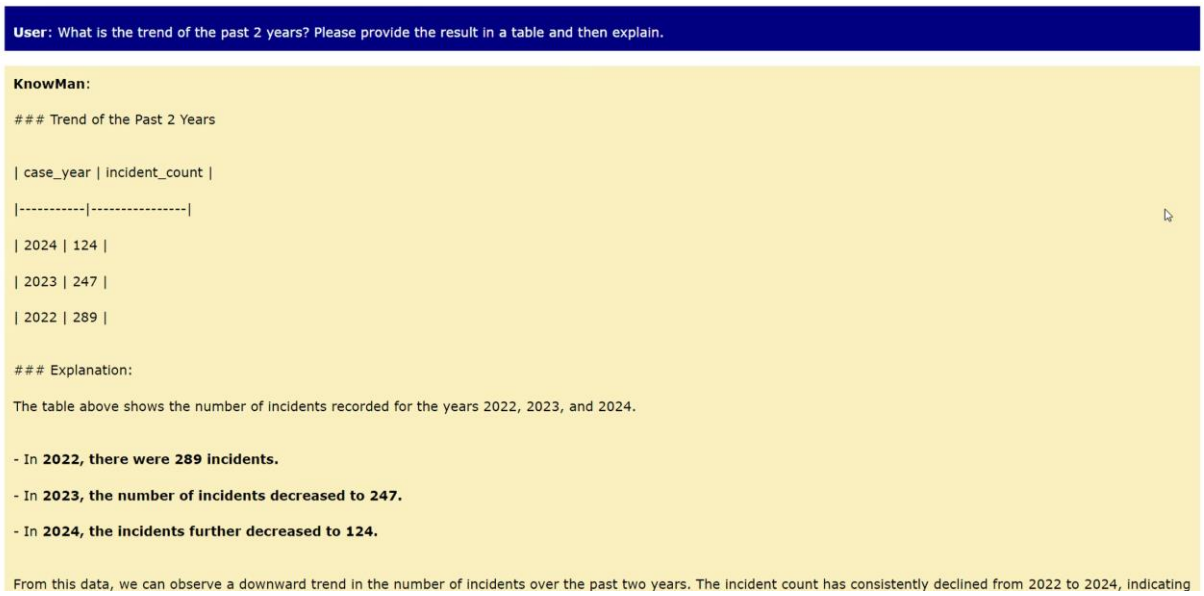
If you ask in Norwegian, the reply is Norwegian.

The model can handle mixed language without problem.

### 6.1.6 Recognising trends




The trend given is simply trend statistics by year. A more specific prompt is needed from the user if another result is desired.



**User**: What is the trend of the past 2 years? Please provide the result in a table and then explain.

**KnowMan**:

### Trend of the Past 2 Years

| case_year | incident_count |
|---|---|
| 2024 | 124 |
| 2023 | 247 |
| 2022 | 289 |

### Explanation:

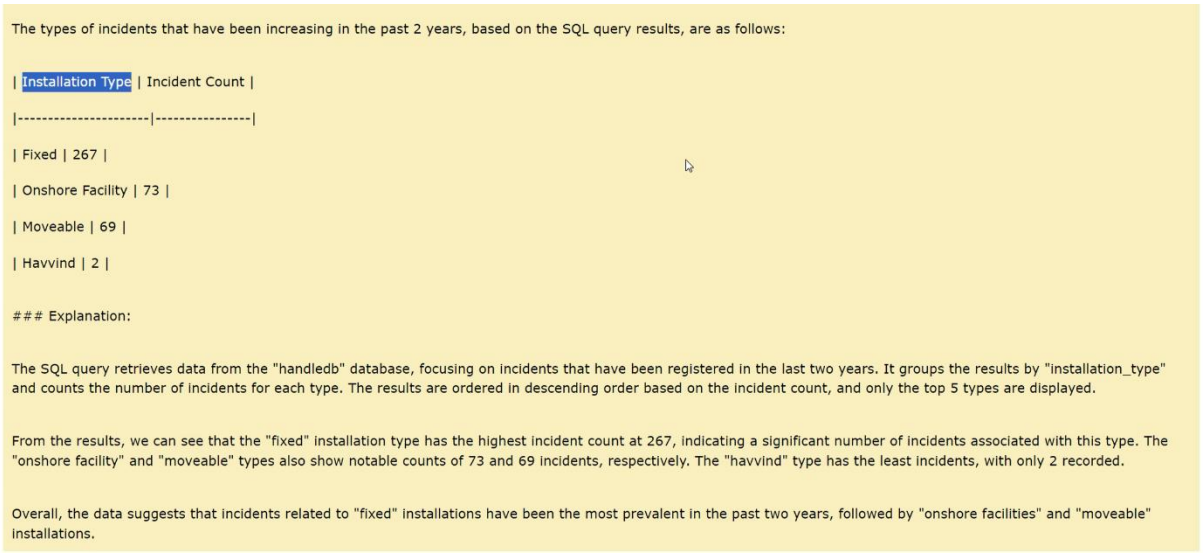
The table above shows the number of incidents recorded for the years 2022, 2023, and 2024.

- In **2022, there were 289 incidents.**
- In **2023, the number of incidents decreased to 247.**
- In **2024, the incidents further decreased to 124.**

From this data, we can observe a downward trend in the number of incidents over the past two years. The incident count has consistently declined from 2022 to 2024, indicating

There is a danger here that the model has not recognised that the year 2024 is not data for the complete year and the trend of decreasing incidents could just be due to partial data for the year 2024.

### 6.1.7 Types of incidents



The types of incidents that have been increasing in the past 2 years, based on the SQL query results, are as follows:

| Installation Type | Incident Count |
|---|---|
| Fixed | 267 |
| Onshore Facility | 73 |
| Moveable | 69 |
| Havvind | 2 |

### Explanation:

The SQL query retrieves data from the "handledb" database, focusing on incidents that have been registered in the last two years. It groups the results by "installation_type" and counts the number of incidents for each type. The results are ordered in descending order based on the incident count, and only the top 5 types are displayed.

From the results, we can see that the "fixed" installation type has the highest incident count at 267, indicating a significant number of incidents associated with this type. The "onshore facility" and "moveable" types also show notable counts of 73 and 69 incidents, respectively. The "havvind" type has the least incidents, with only 2 recorded.

Overall, the data suggests that incidents related to "fixed" installations have been the most prevalent in the past two years, followed by "onshore facilities" and "moveable" installations.

When questioning the model about the types of incidents, the model has used “installation type” Obviously more specificity is needed here.


Asking for incidents related to drilling the model has only returned results where drilling was stopped.

```
SQL code: ```sql
SELECT strftime('%Y', "registered_time") AS "case_year", COUNT(*) AS "incident_count"
FROM handledb
WHERE "drilling_stopped" = 'J' AND "registered_time" >= date('now', '-2 years')
GROUP BY "case_year"
ORDER BY "case_year" DESC
LIMIT 5;
```
```

Using the SQL debug text allowed us to check how the model is designing it's SQL search and here there is some tuning needed such that the model searches the correct column.
```

**Describe incidents that have involved a wrench**

```
SELECT e."description", e."registered_time", p."NameOfPhysicalObject", p."PhysicalObjectNotes" FROM "event" e JOIN "PhysicalObjectDescriptions" p ON e."EventID" = p."EventID" WHERE p."NameOfPhysicalObject" LIKE '%wrench%' LIMIT 5;
```

The incidents that have involved a wrench are as follows:

1. On September 8, 1999, a wrench weighing 250 grams fell down on the board during an event.
2. On September 12, 1999, there was an incident related to the unsatisfactory securing of a pipe wrench during an operation.
3. On May 16, 1999, a pipe wrench weighing approximately 600 grams fell down on the bore deck.
4. On November 21, 1999, a wrench was involved during the assembly of scaffolding at the top of a BOP house.
5. On July 7, 2000, an impact wrench slipped while work was being done on the Coil Tubing deck, resulting in a 46-meter fall.

These incidents highlight various situations where wrenches were involved, including falls and issues with securing tools.

**How many incidents involved an object more than 200 kgs**

```
SELECT COUNT(DISTINCT e."EventID") AS "IncidentCount" FROM "event" e JOIN "PhysicalObjectDescriptions" pod ON e."EventID" = pod."EventID" WHERE pod."PhysicalQuantityValue" > 200;
```

There were 247 incidents that involved an object weighing more than 200 kgs.

**What is the most common incident by object, provide a table.**

```
SELECT "PhysicalObjectDescriptions"."NameOfPhysicalObject", COUNT(*) AS "IncidentCount" FROM "event" JOIN "PhysicalObjectDescriptions" ON "event"."EventID" = "PhysicalObjectDescriptions"."EventID" GROUP BY "PhysicalObjectDescriptions"."NameOfPhysicalObject" ORDER BY "IncidentCount" DESC LIMIT 5;
```

The most common incidents by object are as follows:

| NameOfPhysicalObject | IncidentCount |
|---|---|
| Bolt | 105 |
| Scaffolding plank | 66 |
| Pipe | 51 |
| Drill pipe | 50 |
| Wrench | 42 |

## Om DNV

Vi er et globalt selskap innen kvalitetssikring og risikohåndtering med tilstedeværelse i over 100 land. Vårt formål er å sikre liv, verdier og miljøet. Med vår unike tekniske ekspertise og uavhengighet bistår vi våre kunder med å forbedre sikkerhet, effektivitet og bærekraft.

Enten vi godkjenner et nytt skipsdesign, optimerer energiproduksjonen fra en vindmøllepark, analyserer sensordata fra en gassrørledning eller sertifiserer verdikjeden til en matprodusent, hjelper vi våre kunder med å ta gode og riktige beslutninger og øke tilliten til virksomheten, produktene og tjenestene deres. Verden er i endring. Vi kan påvirke utviklingen. Sammen skal vi takle de globale utfordringene og omstillingene vi vil møte.