



FFI Forsvarets
forskningsinstitutt

21/02647

FFI-RAPPORT

Stordata og avansert analyse

– sluttrapport for FFI-prosjekt «Informasjonsintegrasjon for et moderne forsvar»

Bjørn Jervell Hansen

Jonas Halvorsen

Eirik Anette Flynn Opland

Stordata og avansert analyse
– sluttrapport for FFI-prosjekt «Informasjonsintegrasjon
for et moderne forsvar»

Bjørn Jervell Hansen
Jonas Halvorsen
Eirik Anette Flynn Opland

Emneord

Stordata
Databehandling
Informasjonsinfrastruktur
Informasjonsintegrasjon

FFI-rapport

21/02647

Prosjektnummer

1430

Engelsk tittel

Big data and advanced analytics

Elektronisk ISBN

978-82-464-3384-4

Godkjenner

Jan Erik Voldhaug, *forsknings sjef*
Trude Bloebaum, *forsknings leder*

Dokumentet er elektronisk godkjent og har derfor ikke håndskreven signatur.

Opphavsrett

© Forsvarets forskningsinstitutt (FFI). Publikasjonen kan siteres fritt med kildehenvisning.

Sammendrag

Stordata kan best forstås som et begrep heller enn en konkret teknologi eller familie av teknologier. I dette arbeidet karakteriseres stordata som data av forskjelligartet natur (*variety*), som kommer i store mengder (*volume*) og/eller har hyppig oppdateringsfrekvens (*velocity*) og som et resultat av dette ikke lar seg effektivt håndtere eller bearbeide ved hjelp av tradisjonelle metoder.

Stordataproblemer er, ut fra den samme tankegangen, beregningsoppgaver som enten er så komplekse, vokser så raskt eller fordrer tilgang til så mye data at de ikke kan løses ved hjelp av tradisjonelle metoder. En *stordataløsning* betyr i denne konteksten rett og slett et system som er designet for å løse et stordataproblem.

I en tid der det produseres mer data enn noensinne, vil Forsvaret møte på stordataproblemer når de søker å omsette tilgjengelige data til gode beslutninger. De trenger derfor å øke sin evne til å håndtere disse problemene.

FFI-prosjektet *Informasjonsintegrasjon for et moderne forsvar* har studert stordata og avansert analyse, hovedsakelig gjennom litteraturstudier for å bygge opp kompetanse på sentrale konsepter og teknisk eksperimentering for også å få praktisk erfaring med teknologier forbundet med disse problemstillingene. Formålet med denne rapporten er å oppsummere funnene i prosjektet og kommunisere anbefalinger innenfor fire sentrale områder: Stordataløsninger, kunnskapsgrafer, nevrosymbolisk kunstig intelligens og gjenbruk av modeller.

Teknologier for å håndtere stordataproblemer har lenge vært under enorm utvikling, mye takket være at store aktører deler mange løsningskomponenter av høy kvalitet som åpen kildekode. Dette gjør at tilgangen er stor på tekniske komponenter som kan brukes i stordataløsninger. Vi mener derfor tiden er moden for Forsvaret å ta stordatateknologier i bruk, og anbefaler at:

- Forsvaret setter opp en stordataløsning for et utvalgt stordataproblem for å få erfaring med hvordan slike problemer bør beskrives, hvordan løsninger bør spesifiseres og hvordan løsninger kan realiseres. På den måten vil Forsvaret også få verdifull innsikt i hvilken kompetanse som er nødvendig for å gjennomføre en slik prosess.
- Forsvaret tester bruk av kunnskapsgrafer i et utvalgt analysemiljø for å se om et slikt verktøy gir verdi i form av bedre og raskere analyseresultater.
- Forsvaret setter seg i stand til å ta i bruk teknikker innen nevrosymbolisk kunstig intelligens på et tidlig tidspunkt for automatisert prosessering og analyse av data.
- Forsvaret sikrer seg tilgang til oppdatert kompetanse på gjenbruk av modeller til maskinlæring slik at slike teknikker kan tas raskt i bruk så snart behovet er der.

Dette er tiltak som reduserer risikoen relatert til å anskaffe stordataløsninger, og etter vårt syn vil Forsvaret ved å gjennomføre disse ta et viktig steg i retning av å kunne løse sine nåværende og framtidige stordataproblemer.

Summary

The concept of *big data* remains elusive to define, but is in this work characterized as data of a varied nature (*variety*), that arrives in large amounts (*volume*), and is updated at a high rate (*velocity*), and due to this cannot be efficiently treated with traditional methods.

Using the same starting point, *big data problems* are calculations that cannot be efficiently solved by traditional methods due to the complexity or speed of growth of the task, or the amount of data needed to solve it. *Big data solutions*, correspondingly, are systems designed to solve big data problems.

Today, more data than ever is produced. This means that the Norwegian Armed Forces will encounter big data problems in the process of translating available data into good decisions. They will therefore need to improve their ability to deal with these issues.

The FFI project *Information integration for a modern defense* has studied big data and advanced analysis, mainly through literature studies on key concepts and technical experimentation to also gain practical experience with technologies associated with these issues. The purpose of this report is to summarize the findings of the project and communicate recommendations in four key areas: Big data solutions, knowledge graphs, neurosymbolic artificial intelligence, and model reuse.

The development of technologies for dealing with big data problems have long had a big momentum, thanks in large part to large technology actors sharing many high-quality components as open source. This has led to the situation where the supply of technical components that can be used in big data solutions is plentiful. We therefore believe that it is time for the Norwegian Armed Forces to make use of big data technologies, and recommend that:

- The Armed Forces set up a big data solution for a selected big data problem in order to gain experience of how to describe such problems, and how solutions should be specified and realized. In this way, the Armed Forces will also gain valuable insight into the competence required to carry out such a process.
- The Armed Forces start testing the use of knowledge graphs for a selected analysis task to see if these tools provide value in the form of better and faster analysis results.
- The Armed Forces make sure that they have the ability to use techniques in neurosymbolic artificial intelligence for automated processing and analysis of data at an early stage.
- The Armed Forces ensure access to updated expertise on the reuse of models for machine learning so that such techniques quickly can be taken advantage of as soon as the need arises.

These are measures that will reduce the risk related to acquiring big data solutions, and will in our view ensure that the Norwegian Armed Forces take an important step towards being able to solve their current and future big data problems.

Innhold

Sammendrag	3
Summary	4
1 Introduksjon	7
2 Stordataløsninger	9
2.1 Karakteriserende egenskaper	9
2.2 Typer av stordatakomponenter	11
2.3 Konklusjoner og anbefalinger	13
3 Kunnskapsgrafer	14
3.1 Grafer for integrasjon og analyse	14
3.2 Konklusjoner og anbefalinger	15
4 Nevrosymbolisk kunstig intelligens	17
4.1 Utfordringer og begrensninger med maskinl�ring	17
4.2 Integrasjon av nevr�le og symbolske tiln�rminger	18
4.3 Konklusjoner og anbefalinger	20
5 Gjenbruk av modeller	21
5.1 Eksisterende modeller i ny kontekst	21
5.2 N�r datadistribusjonen endrer seg over tid	21
5.3 Konklusjoner og anbefalinger	22
6 Konklusjon	23
Vedlegg	
Referanser	25



1 Introduksjon

I dagens digitaliserte samfunn blir stadig mer data tilgjengelig på et utall former. Ifølge det amerikanske programvareselskapet Domo, som spesialiserer seg på forretningsinnsikt (*business intelligence*), ble det i 2020 daglig generert 2,5 EB (exabyte)¹ data i verden (Bulao 2021), og det er ingenting som tyder på at denne trenden avtar. Analyseselskapet International Data Corporation (IDC) anslår for eksempel at det totale datavolumet i verden, i det de kaller *the datasphere*, vil øke fra 33 ZB (*zetabyte*)² i 2018 til 175 ZB innen 2025 (Reinsel et al. 2017). En av hovedgrunnene til disse enorme tallene er at Internett gjør det veldig lett å generere og distribuere digitale data: Vi spyr ut 500 000 tweets hver dag (Bulao 2021), Instagram vokser med 95 millioner videoer og bilder hver dag (Lister 2021) og det lastes opp 500 timer med video til YouTube hvert sekund (Wojcicki 2020).

Stordata kan best forstås som et begrep heller enn en konkret teknologi eller familie av teknologier, og det mangler ikke forslag til definisjoner. Vi nøyer oss i denne konteksten med å karakterisere stordata, og gjør dette ved hjelp av det som gjerne kalles de tre V-ene som ble introdusert i Laney (2001): *Volume*, *Velocity* og *Variety*. Vi regner stordata som data av forskjelligartet natur (*Variety*), som kommer i store mengder (*Volume*) og/eller har hyppig oppdateringsfrekvens (*Velocity*) og som et resultat av dette ikke lar seg effektivt håndtere eller bearbeide ved hjelp av tradisjonelle metoder.

Stordataproblemer er, ut fra den samme tankegangen, beregningsoppgaver som enten er så komplekse, vokser så raskt eller fordrer tilgang til så mye data at de ikke kan løses ved hjelp av tradisjonelle metoder. En tilhørende tommelfingerregel er at et problem er et stordataproblem dersom det ikke kan håndteres av én enkelt datamaskin. En *stordataløsning* betyr i denne konteksten rett og slett et system som er designet for å løse et stordataproblem.

Utviklingen av slike løsninger preges av at det nå er en rik tilgang på kostnadsfrie komponenter av god kvalitet, ofte utviklet av kommersielle aktører. De deler sine verktøy åpent og får på den måten gratis hjelp med videreutvikling og feilhåndtering av verktøyene, mens resten av verden får gratis tilgang på programvare som kan gjenbrukes i for eksempel stordataløsninger.

Forsvaret har, som alle andre aktører, et behov for å kunne utnytte sine data til å ta best mulige beslutninger. Dette kommer særlig til syne i målet om at Forsvaret skal ha informasjonsoverlegenhet sammenlignet med sine motstandere (Forsvarsdepartementet 2018). Slik overlegenhet handler om å utnytte alle relevante kilder og bearbeide dataene fra disse for å skape et helhetlig, rettidig og pålitelig situasjonsbilde som gjør at egne styrker har relativt sett bedre situasjonsforståelse enn en potensiell motstander.

Når man ser på en karakterisering av dataene som kreves for å oppfylle denne ambisjonen, ser man at dette kan kalles et stordataproblem:

volume: Forsvaret har tilgang på kilder som potensielt kan levere enorme mengder data, for eksempel Nato Allied Ground Surveillance (AGS) som forventes å levere flere terrabyte (TB³) med data for hver flyvning.

velocity: Forsvaret har tilgang på flere kilder som kontinuerlig leverer data i sanntid, for eksempel

¹ 1 EB \approx 1 000 000 000 000 000 bytes \approx 1 000 000 000 GB.

² 1 ZB \approx 1 000 EB.

³ 1 TB \approx 1 000 GB.

radarsensorer.

variety: Forsvaret har tilgang på mange forskjellige kilder, noe som fordrer at man kan håndtere et mangfold når det kommer til formater, standarder og datamodeller.

Automatisert prosessering og analyse av data har en viktig rolle å spille i å løse stordatap problemer relatert til informasjonsoverlegenhet, som blant annet pekt på i forrige og videreført i gjeldende langtidsplan for Forsvaret (Forsvarsdepartementet 2016, Forsvarsdepartementet 2020). Dette skyldes at om man skal kunne omsette de store datamengdene til grunnlag for beslutninger kan man ikke basere seg på kun manuell behandling, men må finne måter å utnytte de stadig økende mulighetene som moderne informasjonsteknologi gir.

Hvordan Forsvaret skal kunne øke sin evne til automatisert prosessering og analyse samt håndtering av stordata, var en sentral problemstilling da FFI-prosjektet *Informasjonsintegrasjon for et moderne forsvar* ble startet i 2017. Prosjektets aktivitet har i hovedsak bestått i litteraturstudier for å bygge opp kompetanse på sentrale konsepter, samt teknisk eksperimentering for også å få praktisk erfaring med teknologier og teknikker forbundet med disse problemstillingene. Prosjektets resultater kan deles inn i fire hoveddeler, og denne rapporten er organisert etter disse:

- Stordataløsninger, presentert i kapittel 2.
- Kunnskapsgrafer, presentert i kapittel 3.
- Nevrosymbolsk kunstig intelligens, presentert i kapittel 4.
- Gjenbruk av modeller, presentert i kapittel 5.

Funnene blir oppsummert og rapporten konkludert i kapittel 6.

Siden vi her omtaler teknologier designet for å løse viktige utfordringer for Forsvaret, mener vi at våre funn bør tas i betraktning når det skal anskaffes stordataløsninger. Rapporten henvender seg derfor først og fremst til personell med ansvar for å spesifisere og vurdere slike anskaffelser til Forsvaret.

2 Stordataløsninger

Med utgangspunktet at et stordataproblem ikke lar seg løse av én enkelt datamaskin, er det åpenbart at stordataløsninger i hovedsak krever flere datamaskiner og derfor er systemer som utnytter parallell og distribuert prosessering. Slike systemer finnes det mye erfaring med innen fagfeltet, og dette danner mye av grunnlaget for moderne stordataløsninger (Senger & Geyer 2016).

Å bygge en god stordataløsning består i å sette sammen programvarekomponenter som til sammen gir de egenskapene som kreves for å løse det konkrete stordataprobblet man har. Det viser seg at å designe en generell løsning for stordataprobler, slik man i sin tid kunne ty til relasjonsdatabaser for å løse sine datautfordringer, ikke er mulig. Det eneste som er felles for forskjellige stordataløsninger er at de er distribuerte.

Vår erfaring gjennom dette prosjektet er at det er rik tilgang på kostnadsfrie komponenter av god kvalitet som kan settes sammen til kraftige stordataløsninger, noe våre eksperimenter beskrevet i Stolpe et al. (2020), Halvorsen & Hansen (2020) og Hansen (2021) illustrerer. Dette betyr at dersom man har tilgang til riktig kompetanse kan man selv sette sammen slike løsninger heller enn å kjøpe ferdig sammensatte løsninger i markedet. Uansett om en løsning kjøpes som ett system eller om man selv setter den sammen, er det viktig for den som har behov for en stordataløsning å være i stand til å beskrive sitt stordataproblem i en slik detalj at man kan identifisere hvilke egenskaper løsningen må ha. Dette er ingen triviell oppgave, men krever en kombinasjon av god forståelse av problemet som skal løses og teknisk kunnskap om egenskaper ved potensielle komponenter.

I prosjektet som denne rapporten oppsummerer har vi studert hvilke egenskaper som er de viktigste, og i det følgende vil vi presentere en liste av egenskaper vi mener har stor betydning for oppførselen og anvendeligheten av en stordataløsning og som derfor er viktige å ta med i betraktningen når slike skal utformes.

Etter en gjennomgang av disse egenskapene presenterer vi en inndeling av komponenter som kan inngå i en stordataløsning. Til slutt oppsummerer vi våre erfaringer fra denne delen av prosjektet og gir anbefalinger.

2.1 Karakteriserende egenskaper

Hvilke egenskaper som er nødvendige ved en stordataløsning er altså tett knyttet til problemet som skal løses. Det betyr at vidt forskjellige problemer kan forventes å kreve helt forskjellige egenskaper av løsningen.

Aktivitetene i prosjektet ble derfor designet med det utgangspunktet at jo flere typer problemer vi studerer, jo flere viktige egenskaper vil vi oppdage. Vi har imidlertid verken hatt ressurser til, eller hatt mål om, å dekke alle mulige typer problemer, så samlingen av egenskaper vi presenterer kan ikke forventes å være komplett.

Denne lista over egenskaper kan brukes både til å karakterisere problemstillingen og til å beskrive potensielle løsningskomponenter, og kan som sådan bidra til å dekke gapet mellom problemstilling og teknologi. Dersom man for eksempel har en kontinuerlig strøm av data man ønsker å overvåke i sanntid, kreves det av løsningen at den kan produsere nødvendige resultater i løpet av få sekunder.

Komponenter som ikke kan oppfylle dette kravet vil da neppe være til særlig nytte for å løse et stordataproblemet med dette som en viktig forutsetning.

Egenskapene vi har identifisert i vårt arbeid er:

- Programmeringsmodell
- Skalerbarhet
- Tilgjengelighet eller konsistens
- Feiltoleranse
- Responsivitet
- Forklarbarhet
- Fleksibilitet

Disse vil bli forklart i det følgende.

Hvilken **programmeringsmodell** er best egnet for å løse det aktuelle stordataproblemet? Med en programmeringsmodell menes her den grunnleggende programmeringsstilen som en komponent eller et rammeverk legger opp til at utviklere skal uttrykke seg i. Forskjellige programmeringsmodeller har forskjellige styrker og svakheter som bør settes opp mot de ønskede egenskapene til løsningen. Eksempler på mye brukte programmeringsmodeller er funksjonell programmering og aktørmodellen (*actor model*), som begge er velegnede for å støtte parallell og distribuert prosessering.

Hva slags **skalerbarhet** trenger løsningen for å kunne løse det aktuelle stordataproblemet også over tid? Skalerbarhet kan defineres som evnen til å håndtere en økende mengde arbeid, eller evnen systemet selv har til å imøtekomme en slik økning. Flere forskjellige former for skalerbarhet kan være viktig, avhengig av situasjon. Teknisk skalerbarhet (skalerbarhet i datamengde og regnekraft) og funksjonell skalerbarhet (hvor lett det er å legge til ny funksjonalitet) er to eksempler.

Hva er viktigst for løsningen: At resultater alltid er **tilgjengelig** overalt, eller at resultatene, når de først er tilgjengelige, er like i hele systemet, altså **konsistente**? Siden stordataløsninger i hovedsak er distribuerte systemer, vil det såkalte CAP-teoremet (Gilbert & Lynch 2002) slå inn: I et distribuert system kan man ikke garantere både konsistens (*consistency* – C) og tilgjengelighet (*availability* – A) dersom man samtidig ønsker at systemet skal fungere som normalt også dersom deler av det distribuerte systemet ikke kan nås (*partition tolerance* – P). Som et eksempel kan man tenke seg at det i et distribuert system som skal finne målkoordinater som skal brukes til å lede ild er viktigere at koordinatene er korrekte og de samme i hele systemet enn at de er tilgjengelige til enhver tid, og at en løsning her burde prioritere konsistens over tilgjengelighet.

Hvordan skal løsningen oppnå ønsket **feiltoleranse**? Et system er feiltolerant dersom det kan velge å degradere framfor å avslutte dersom én eller flere komponenter svikter. Degradering kan her for eksempel bety at data går tapt, men at systemet fortsetter som normalt med dataene som er tilgjengelig, at responstiden blir lenger eller at systemet holder inkonsistente data. Feiltoleranse kan løses på forskjellige måter, og er ofte tett forbundet med selve arkitekturen i løsningen. To eksempler er *master/slave*-arkitektur, der enkelte noder (*master*) i løsningen tar ansvaret for feilsituasjoner og dikterer hvordan de andre nodene (*slave*) skal oppføre seg, og *peer-to-peer*-arkitektur, der nodene i løsningen koordinerer arbeidsfordeling seg imellom. Komponenter som velges til en stordataløsning må kunne håndtere den ønskede feiltoleransestrategien.

Hvor **responsiv** må løsningen være; hvor raskt må den kunne levere resultater? Må resultatene for

eksempel leveres i sanntid for å være nyttige? Systemer som skal levere resultater i sanntid eller nær sanntid, må i så fall ha komponenter som kan prosessere data tilstrekkelig raskt. Det betyr for alle praktiske formål at komponentene i systemet må være primærminnekomponenter, altså komponenter som kjører i datamaskinens minne (*in-memory*) og som ikke er avhengig å gjennomføre tidkrevende lese- og skriveoperasjoner mot persistente datalager.

I hvilken grad er **forklarbarhet** viktig? Hvor viktig er det at en bruker av løsningen vet hvordan resultatene er blitt til? Dette er en egenskap som er spesielt viktig i en militær sammenheng der stordataløsninger i bunn og grunn etableres for å bidra til beslutningsstøtte for militære beslutningstakere. I tilfeller der en stordataløsning skal underbygge beslutninger som potensielt kan handle om liv og død, er det rimelig at en beslutningstaker forventer å forstå grunnlaget for konklusjoner som trekkes. Dette er spesielt aktuelt når det gjelder stordataløsninger som inkluderer bruk av kunstig intelligens, og både norske myndigheter og NATO setter krav til forklarbarhet i slike tilfeller (Regjeringen Solberg 2020, NATO Ministers of Defence 2021).

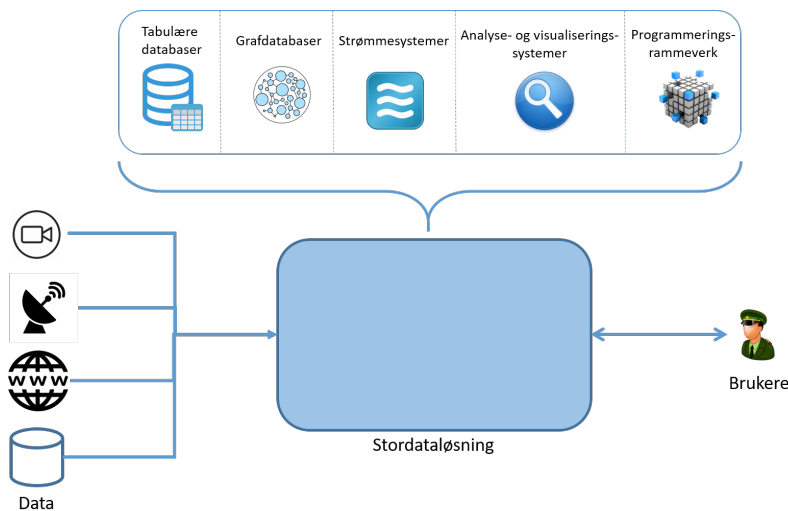
Hvor **fleksibel** må løsningen være? Kan det forventes at løsningen må kunne håndtere nye datakilder eller nye løsningskomponenter mens den er i drift? Dersom dette er ønskelig må løsningen ha en arkitektur som legger til rette for løst koblede komponenter, og komponentene som velges i en slik løsning må være i stand til å automatisk tilpasse seg nye komponenter å virke sammen med.

2.2 Typer av stordatakomponenter

Utviklingen av stordatakomponenter beveger seg svært hurtig, noe som både gir et stadig tilsig av nye komponenter, men som også gir en usikkerhet rundt hvilke komponenter som satses på og videreutvikles. Dette forsterkes av at mye av utviklingen skjer i form av åpen kildekode (*open source*): Det mangler ofte kommersielle insentiver til å vedlikeholde komponenter, noe som kan gjøre at utviklere forlater prosjektet og heller starter noe nytt. Denne ustabiliteten i løsningskomponenter kan gi utfordringer når det kommer til drift og vedlikehold av stordataløsninger. På grunn av denne store utviklingshastigheten har vi i prosjektet ikke hatt som mål å identifisere hvilke komponenter som per dags dato er de beste innenfor sine felt, men heller studert komponenter som representanter for komponenttyper.

Vi presenterer derfor her en inndeling av komponenter i fem grove hovedkategorier. Denne inndelingen er én av flere mulige, og hovedtypene som presenteres er delvis overlappende. Vi tror allikevel den gir et nyttig bilde av hva slags komponenter man skal se etter dersom man ønsker å lage en løsning til sitt stordataproblem. Dette bør imidlertid ikke tolkes dit hen at man må velge minst én komponent fra hver av disse hovedtypene. Hvilke komponenter som kreves i en stordataløsning er avhengig av hvilket stordataproblem som skal løses. Hovedtypene av komponenter er (se også figur 2.1):

- Tabulære databaser
- Grafdatabaser
- Strømmesystemer
- Analyse- og visualiseringssystemer
- Programmeringsrammeverk



Figur 2.1 En stordataløsning vil typisk inneholde én eller flere av de viste komponenttypene.

Tabulære databaser forstås fortsatt oftest som relasjonsdatabaser, men i kontekst av stordata er det først og fremst de såkalte NoSQL-databasene⁴ som er mest brukt. Tabulære databaser representerer data i forskjellige former for tabeller, og i en stordatakontekst er dette ofte enkle tabeller der hvert objekt (*value*) er identifisert med en nøkkel (*key*) i henhold til det velkjente *key-value*-paradigmet (Meier & Kaufmann 2019).

Grafdatabaser inkluderes som regel i begrepet NoSQL-databaser, men har egenskaper som gjør at vi her fremhever disse som en egen type. Dette er databaser der dataene holdes i graf-form: de representeres som en samling datapunkter som er koblet sammen.⁵ En slik representasjon er velegnet for å støtte analyse av data der relasjoner er spesielt viktige, for eksempel data om sosiale nettverk (Brandes et al. 2014). Flere detaljer om grafer og hvordan de kan utnyttes til å håndtere stordata finnes i kapittel 3.

Strømmesystemer er systemer som er designet for å behandle vilkårlig lange sekvenser av sanntidsdata. Slike systemer legger til rette for fortløpende analyse av dataene og rask levering av resultater (helst i sanntid). Dette oppnås som regel ved at dataene holdes i systemets primærminne (RAM – *Random Access Memory*).

Analyse- og visualiseringssystemer er systemer designet for å hjelpe brukere til å forstå betydningen av dataene ved å sette dem i en visuell sammenheng. Komponenter av denne typen er ofte inkludert når man skal designe stordataløsninger som skal støtte visuell analyse av data.

Programmeringsrammeverk er et ferdigpakket sett av programvarebiblioteker og lagringsløsninger som ved hjelp av brukerkode kan tilpasses spesifikke problemer og behov. Slike rammeverk inkluderer som oftest komponenter fra en eller flere av de andre komponenttypene, men framheves her som en egen type fordi det innenfor stordatafeltet er flere velkjente slike løsninger.⁶

⁴Forkortelsen NoSQL leses ofte som *not only SQL* og kan forstås som en samlebetegnelse for alle databaser som ikke begrenser seg til relasjonsmodellen som ligger til grunn for relasjonsdatabaser.

⁵På fagspråket kalles disse *noder* (datapunkter) og *kanter* (koblinger).

⁶For eksempel Apache Hadoop (<https://hadoop.apache.org/>) og Apache Spark (<https://spark.apache.org/>).

2.3 Konklusjoner og anbefalinger

Å lage løsninger for stordatap problemer er en omstendelig prosess som fordrer at problemstilling, tilgjengelige data og mulige tekniske løsninger må analyseres og forstås godt før en faktisk løsning kan etableres. Dette krever en grundig forståelse av disse aspektene, og fordrer et samarbeid mellom domeneeksperter og teknologer. Effektive stordataløsninger krever skreddersøm basert på problemstilling, og det eneste som virker å være felles for slike løsninger er at de i sin natur er distribuerte.

Det er god tilgjengelighet på kostnadsfrie komponenter av høy kvalitet som er klare til å kunne brukes i stordataløsninger. Vår erfaring er imidlertid at det kreves god innsikt både i generelle IKT-prinsipper og de spesifikke komponentene for å sette sammen gode fleksible og robuste løsninger. Slike tilpassede løsninger virker også å være ressurskrevende å drifte og vedlikeholde på grunn av den høye utviklingstakten innenfor feltet.

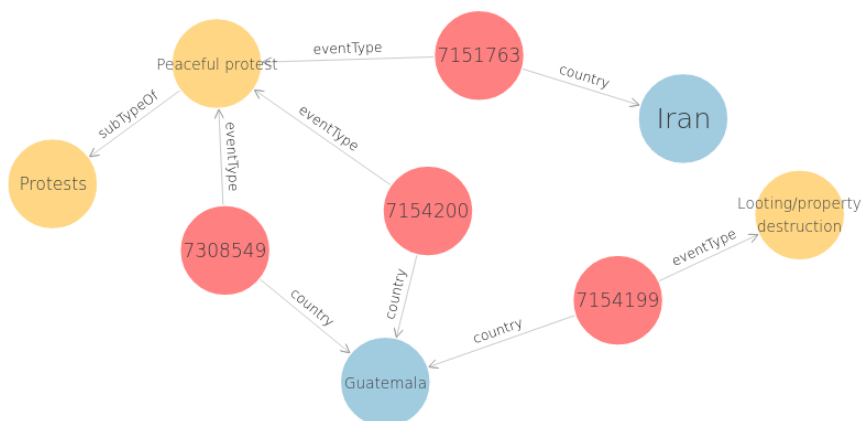
Med utviklingen man ser innen dataproduksjon og behovet for å prosessere dataene for å kunne ta gode beslutninger, er det rimelig å forvente at behovet for stordataløsninger vil øke også i Forsvaret. Stordataløsninger er mye brukt i det sivile, men for å designe løsninger som tar hensyn til forsvarsspesifikke behov trenger Forsvaret erfaring med hvordan slike løsninger kan spesifiseres og realiseres. Med dette som utgangspunkt anbefaler vi at Forsvaret setter opp en stordataløsning for et utvalgt stordataproblem. På den måten vil Forsvaret få verdifull innsikt i hvilken kompetanse som er nødvendig for å gjennomføre en slik prosess. For å sikre god forankring til problemet som skal løses bør arbeidet ledes av miljøet i Forsvaret som eier den aktuelle problemstillingen, men det følger ikke nødvendigvis av dette at det er Forsvaret selv som også skal gjennomføre oppsett av den tekniske løsningen.

3 Kunnskapsgrafer

Kunnskapsgrafer baserer seg på prinsippet om å representere data i form av grafer. Grafer er svært versatile og fleksible datastrukturer, og mange typer av problemer lar seg naturlig formulere som abstrakte grafproblemer (Stolpe et al. 2019, kap. 3.2).

En graf består av noder og kanter som fanger relasjoner mellom dataentiteter, som illustrert i figur 3.1. For eksempel er veldig mye forskjellig data relasjonell av natur og lar seg naturlig representeres som grafer. Intuitive eksempler på relasjonell data er molekylære strukturer (for eksempel for legemiddelforskning), sosiale nettverk og trafikflyt. En kunnskapsgraf kan sådan beskrives å være en graf bestående av forskjelligartet informasjon.

Kunnskapsgrafer brukes i omfattende grad til problemstillinger der man trenger å integrere og trekke ut verdi fra varierende heterogene kilder med potensielt mye data (Hogan et al. 2021). Det mest kjente eksempelet på en kunnskapsgraf er *Google Knowledge Graph*⁷. Det finnes per i dag ingen entydig definisjonen på hva som utgjør en kunnskapsgraf, men i denne rapporten følger vi en ganske vid definisjon fra Hogan et al. (2021); *en kunnskapsgraf er en datagraf som uttrykker kunnskap om den virkelige verden, hvor noder representerer entiteter av interesse og kanter relasjonene mellom disse entitetene*. Vi vil i det følgende bruke begrepene kunnskapsgraf og graf om hverandre.



Figur 3.1 Eksempel på en kunnskapsgraf. De forskjellige fargene representerer forskjellige typer data, i dette tilfelle er røde noder hendelser, gule noder hendelsestyper og blå noder geografiske steder.

3.1 Grafer for integrasjon og analyse

Som nevnt er grafer en velegnet datastruktur for å integrere og navigere i forskjelligartet informasjon. Grafer er derfor spesielt godt egnet for typisk flerkilde analyseoppgaver hvor man leter etter sammenhenger over data fra flere kilder.

⁷ Brukes bl.a. for å lage informasjonsboksen som ofte dukker opp til høyre i bildet når man får svar fra et Google-søk.

Grafer er i utgangspunktet skjematøse⁸, og setter få begrensninger på representasjon annet enn at informasjonen må kunne representeres som kanter og noder. For å gjøre det enda lettere å sammenstille data, kan man ta i bruk prinsipper som unike identifikatorer og gjenbrukbare skjema.

Å forene to grafer er en triviell operasjon, da unionen av to grafer er en graf, men uten bruk av felles identifikatorer vil informasjonen ikke umiddelbart være sammenstilt. Dette får man lettere til hvis man publiserer data som *lenkede data* (Heath & Bizer 2011, Direktoratet for forvaltning og IKT 2019). Lenkede data er en strategi for hvordan man kan representere data for å forenkle gjenbruk og integrasjon av forskjelligartede datasett, og baserer seg på å bruke standarden *Resource Description Framework* (Cyganiak et al. 2014) fra World Wide Web Consortium (W3C) for å representere data som grafer, samt unike identifikatorer for kanter og noder. Disse identifikatorene formes som regel som web-adresser (URI⁹), noe som gjør det lett å knytte data sammen med andre data.

Som nevnt ovenfor er grafer i utgangspunktet skjematøse, noe som gjør dem fleksible og godt egnet til å fange ufullstendig informasjon. For øvrig kan man ta i bruk skjemaer ved behov, på enten hele eller deler av grafen, for enten å øke gjenbrukbarheten ved å bruke velkjente begrepsmodeller, validere data, eller å utlede implisitt informasjon ved bruk av formelle *ontologier* (Stolpe et al. 2016, Hogan et al. 2021) som tilegner semantikk til nodene og kantene. Kort sagt, så vil skjemaer gjøre det lettere for brukere å finne og utnytte informasjon.

Når det kommer til selve analyseoppgaven, er det flere typer analyser man kan tenkes å gjøre. En veldig naturlig oppgave som er direkte tilgjengelig ved at data er i graf-form er *traversering* av grafen, for eksempel for å identifisere interessante stier i et kontaktnettverk. Her kan man applisere graf-algoritmer som er velkjent fra sosial nettverksanalyse, som identifisering av klynger og nøkkelpersoner i nettverket (sentralitet) for å gi økt innsikt (De Nooy et al. 2018). Videre kan man applisere logikkbasert resonnering i form av regler og ontologier for å utlede implisitt informasjon, som for eksempel klassifisering (Stolpe et al. 2016). I tillegg kan man applisere dyplæring på grafer i form av *Graph Neural Networks* som kan predikere både klassifisering av noder og kanter mellom noder i de situasjonene hvor data er ufullstendig (Hogan et al. 2021).

Kunnskapsgrafer, og assosierte teknologier og standarder, er sådan velegnet for oppgaven med å integrere og hente ut verdi fra informasjon spredt over heterogene datasett.

3.2 Konklusjoner og anbefalinger

Kunnskapsgrafer er kraftige og fleksible verktøy for å representere, sammenstille og analysere informasjon fra forskjellige kilder og dermed bidra til å løse stordatap problemer innenfor typen *variety*. Vår erfaring er at også strømmende data (*velocity*) kan håndteres i graf-format, se for eksempel (Halvorsen & Hansen 2020). På grunn av fleksibiliteten i graf-formatet, legger grafer også godt til rette for gjenbruk av data, spesielt når man i tillegg utnytter konseptene rundt lenkede data (Heath & Bizer 2011), som unike identifikatorer og gjenbruk av vokabularuttrykk (Halvorsen & Hansen 2020).

Det er god tilgang på modne og robuste produkter innenfor dette feltet, både gjennom åpen kildekode

⁸Skjema er en modell som definerer dataene på et konseptuelt nivå.

⁹Uniform Resource Identifier.

og kommersielle produkter, som kan utnyttes for å ta kunnskapsgrafer i bruk i egnede domener i Forsvaret. Vår erfaring tilsier at analyseaktiviteter er blant områdene der kunnskapsgrafer og grafbaserte verktøy burde kunne tas i bruk umiddelbart og gi gevinst i form av bedre og raskere analyseresultater.

På grunn av den relativt gode tilgangen på løsninger innenfor dette feltet, anbefaler vi at Forsvaret tester en slik løsning i et domene der man kan forvente at en graf-tilnærming kan gi rask gevinst. På den måten vil man kunne teste hypotesen om at kunnskapsgrafer er velegnet også i en militær kontekst der det for eksempel er mye strengere krav til finkornet tilgangskontroll enn det som typisk er støttet i kommersielle produkter. Med tanke på egenskapene til en kunnskapsgraf, virker et analysemiljø å være et godt sted å gjøre en slik uttesting.

4 Nevrosymbolisk kunstig intelligens

Maskinlæring, derav spesielt nevrale nett (NN) og dyplæring (DL), er områder innen kunstig intelligens (AI — *Artificial Intelligence*) som har hatt en eventyrlig vekst de siste årene. AI kan vise til imponerende resultater for eksempel når det kommer til bilde- og tekstforståelse (Krizhevsky et al. 2012, Mikolov et al. 2013), proteinfolding (Jumper et al. 2021) samt brettspill som sjakk og Go (Silver et al. 2018). Med den kraftig økende mengden med sensorer blir automatisering av analyseoppgaver mer og mer nødvendig for å redusere informasjonsoverlast i en forsvarssammenheng (Reding & Eaton 2020), og det er lett å tro at NN/DL slik det brukes i sivil sektor umiddelbart kan brukes til slike oppgaver.

Disse teknologiene har dog visse iboende og grunnleggende begrensninger som kan gjøre det problematisk å ta dem i bruk i forsvarssammenheng. Spesielt mangelen på gjennomsiktighet, innsikt og forståelse om *hvorfor* et nevralt nett konkluderer slik det gjør, har direkte innvirkning på tilliten man kan ha til resultatet. Disse begrensningene kan være uakseptable for en del bruksområder, spesielt når resultatet er en del av grunnlaget for å ta beslutninger som kan ha alvorlige konsekvenser. En analytiker vil sannsynligvis trenge innsikt i hvilke data som har hatt betydning for resultatet og deres relative vektlegging for å kunne forstå utkommet og bedømme dens kvalitet, og jo mer uforståelig resultatene er jo mer må analytikeren bruke tid på å krysssjekke og ettergå resultatene manuelt (Gunning & Aha 2019).

For øvrig er det føringer fra både NATO og nasjonale myndigheter som tilsier at AI skal være forklarbart. Regjeringen Solberg la frem en nasjonal strategi for AI, der man sier at Norge skal sikte på å operere i henhold til prinsippene ansvarlig og pålitelig AI samt legger seg på prinsippene og retningslinjene som EU-kommisjonen har kommet opp med (European Commission 2019) Spesifikt nevnes det at “Beslutninger tatt av systemer basert på AI, skal være sporbare, forklarbare og gjennomsiktige” (Regjeringen Solberg 2020). NATO vedtok på ministermøtet 21.–22. oktober 2021 en strategi for AI, som spesifiserer at alle AI-applikasjoner som utvikles av allierte blant annet skal være i henhold til prinsippene for forklarbarhet, sporbarhet og ansvarlighet (NATO Ministers of Defence 2021). Eksempelvis skal AI-applikasjonene være forståbare og transparente.

Føringene er nokså tydelige; forklarbarhet, gjennomsiktighet, og ansvarlighet i AI blir essensielt fremover. Et aktivt forskningsfelt som prøver å adressere disse overnevnte utfordringene er feltet *nevro-symbolisk* AI, som kombinerer nevrale nett med formell logisk resonnering.

4.1 Utfordringer og begrensninger med maskinlæring

Det mest presserende problemet med nevrale nett og dyplæring for militært bruk er altså mangel på *forklarbarhet*. Et trent nevralt nett blir for alle praktiske formål en svart boks hvor man ikke har innsikt i hvorfor systemet har nådd de konklusjonene det har produsert. Man kan ikke uten videre titte inn i det nevrale nettet og få en konseptuell forståelig forklaring eller intuisjon på hvorfor resultatet er som det er; nettverket består av noder med trente vektter og terskelverdier som det har tilegnet seg basert på et datasett.

For situasjoner hvor beslutninger som må tas kan ha alvorlige konsekvenser, kan det å lene seg på resultater hvor innsikt i resonnement er fraværende være uakseptabelt både fra et pragmatisk,

etisk og juridisk perspektiv. Ofte vil det være et krav om at beslutninger må begrunnes nøye, og informasjonsgrunnlaget som beslutningene tas på må kunne vurderes, veies og senere ettergås. Mangelen av forklarbarhet kan derfor forhindre militær bruk av NN/DL-mulighetene som er i omfattende bruk i sivil sektor per i dag.

Andre velkjente begrensninger med NN/DL er 1) avhengigheten av store mengder med treningsdata, noe som ikke alltid er mulig å oppdrive, 2) bias som følge av overtrening, 3) sårbarhet for fiendtlig manipulering (*adversarial attack*) (Garcez & Lamb 2020) og 4) at NN/DL er dårlig egnet til å ta i bruk (eller la seg kombinere med) etablert kunnskap. Når man kommer til analyseoppgaver som er komplekse og komposisjonelle av natur, og med lav hendelsesfrekvens, som deteksjon av høynivåhendelser som spres seg over tid og rom (eksempelvis det å forstå en aktørs oppførsel over tid), er det mye som tyder på at NN/DL alene ikke er velegnet som løsning.

4.2 Integrasjon av nevralt og symbolske tilnæringer

Den siste tiden har vi sett økt fokus på å kombinere nevralt tilnæringer med *symbolske* tilnæringer for å adressere blant annet utfordringen med forklarbarhet. Nevrale nett og dyplæring tilhører den større gruppen AI-teknologier som refereres til som *subsymbolske* tilnæringer, og tradisjonelt utgjør de to separate gruppene subsymbolisk og symbolsk samlet sett fagfeltet AI (se figur 4.1), hvor førstnevnte er basert på statistiske og biologisk inspirerte modeller og sistnevnte på logisk resonnering. Disse fagfeltene har historisk sett hatt lite kryssinteraksjon, men man har de siste årene sett et tydelig behov for å kombinere de to for å oppnå de samlede fordelene de vil kunne tilby. Dette kombinerte fagfeltet går under tittelen *nevro-symbolisk AI*, og akademiske eksperter spår at det er innenfor dette subfeltet av AI man vil se størst fremskritt i det neste tiåret (Littman et al. 2021).

Daniel Kahneman, en professor i psykologi og nobelprisvinner i økonomi, skisserte i sin velkjente bok “Thinking fast and slow” (Kahneman 2011) teorien om at mennesker har to grunnleggende systemer for tenkning som styrer beslutninger:

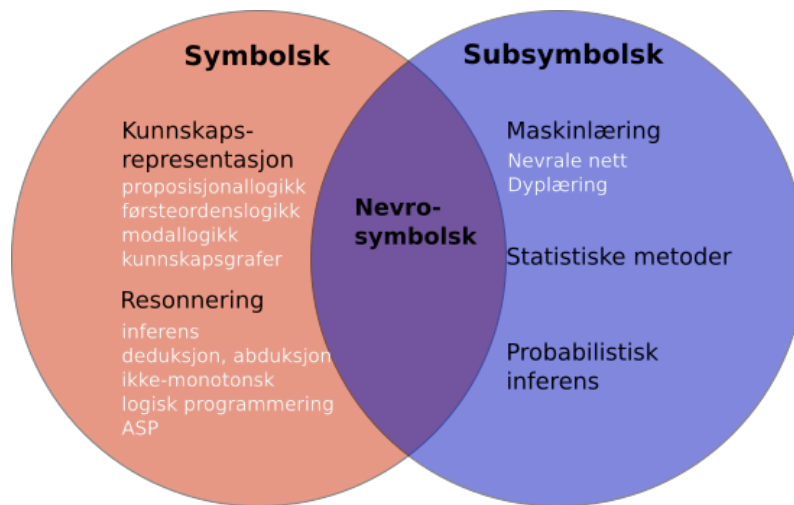
System 1: Thinking Fast Hurtig, instinktiv, ubevisst, upresis, sensitiv for bias

System 2: Thinking Slow Sakte, logisk, bevisst og rasjonell

Disse to menneskelige systemene er nokså analoge til henholdsvis maskinlæring og logisk resonnering, hvor System 1 kan tilskrives NN/DL og System 2 den symbolske delen av AI. Det er derfor ikke unaturlig å tenke seg at kombinasjonen av disse to tilnærmingene vil vise seg å være nødvendig for å oppnå ønskede egenskaper (Lex Fridman 2020).

Fordelene med subsymbolske teknologier er mange, hvor særlig læring fra data har vist seg å være spesielt effektivt i praksis. Symbolske teknologier, som bruker formalisert kunnskap, er dog bedre egnet når man har lite datagrunnlag, men god tilgang til domenekunnskap og/eller forklarbarhet eller logisk korrekthet er essensielt.

Det første tilfellet kan for eksempel manifestere seg i typisk analyseoppgaver; handlingsmønsteret til en aktør kan endre seg hyppig, og en domeneekspert kan lettere identifisere en kritisk, men tilsynelatende enkel, endring i oppførselsmønster selv med lite datagrunnlag. Det vil være veldig gunstig å kunne utnytte formalisert domenekunnskap algoritmisk, noe som NN/DL-tilnæringer



Figur 4.1 Oversikt over felter innen AI.

ikke er godt egnet til.¹⁰ Det andre tilfellet adresseres via symbolske tilnæringer ved at forklarbarhet er innebygget i disse tilnærmingene fra bunnen av i form av strukturert kunnskap, deduktive regler og logisk resonnering. Man er i slike tilfeller derfor i stand til å utlede resonneringer som kan ettergås av en beslutningstaker (også kjent som *eXplainability*). Svakheterne med symbolske tilnæringer er at det er avhengig av strukturert kunnskap, i motsetning til NN/DL som kan lære fra ustrukturerte rådata og utlede informasjon fra det.

Det er mange måter å kombinere disse to tilnærmingene, fra å la en symbolsk resonneringsmotor kunne utnytte NN/DL på atomære oppgaver (som for eksempel detektere objekter i bilder), til det å utføre logisk resonnering på nodene i et nevralt nett (Kautz 2020). Dette er et aktivt og pågående forskningsfelt som per i dag er på et lavt modenhetsnivå.

Behovet for neurosymboliske teknologier for militært bruk er imidlertid allerede veletablert i litteraturen. Eksempelvis finner vi neurosymboliske teknologier brukt for å adressere forklarbarhet og tillit i data delt i militære koalisjoner (Preece et al. 2021), ruting av konvoier i forbindelse med kompliserte angrepsscenarioer (Xing et al. 2019) samt for å detektere og skille mellom forskjellige typer IED-hendelser (*Improvised Explosive Devices*, Preece et al. (2021)). Følgende sitat fra army.gov oppsummer hvorfor en neurosymbolisk tilnærming er nødvendig for hendelseshåndtering:

«*The dimensionality of the sensor data whether it is time-series data (e.g. accelerometers) or video is infeasible for pure symbolic reasoning. Similarly, deep learning is unable to learn patterns that manifest over large time and spatial scales inherent in complex events. The hybrid of neuro-symbolic learning is necessary for complex event processing.*» (U.S. Army DEVCOM Army Research Laboratory Public Affairs 2020)

¹⁰Denne egenskapen refereres noen ganger til som *tellability* i litteraturen.

4.3 Konklusjoner og anbefalinger

Bruk av maskinlæring og dyplæring på oppgaver der utfallet kan ha store menneskelige konsekvenser kan være problematisk på grunn av mangel på innsyn og forståelse av underliggende modell. I tillegg spesifiserer nasjonale og organisatoriske føringer at AI skal være forklarbart (Regjeringen Solberg 2020, NATO Ministers of Defence 2021). Nevrosymbolske tilnærminger ser ut til å kunne adressere dette aspektet, og er per i dag et meget aktivt forskningsfelt der systemer foreligger på et relativt lavt modenhetsnivå.¹¹

De store fremskrittene innen dyplæring de siste tiårene har kommet fra de store teknologiaktørene i sivil sektor. Det virker derimot litt mindre sikkert at disse aktørene vil være like ledende når det kommer til nevrosymbolisk AI; det er tross alt oppgaver som for eksempel profilering av reklame som har vært hoveddriver for den teknologiske utviklingen fra disse aktørene. Forklarbarhet, gjennomsiktighet og ansvarlighet er trolig mye viktigere faktorer for militære anvendelser, der avgjørelser kan ha alvorlige konsekvenser, enn det er for reklamesalg.

Når det er sagt, så er dette et aktivt forskningsfelt i akademien som har viktige bruksområder i sivil sektor; forklarbarhet er åpenbart nyttig i medisinsk bruk av NN/DL, samt at EU-lovgivning (i prinsippet) krevet at man som individ skal kunne få innsyn i logikken som blir brukt av AI når avgjørelser blir utledet basert på personinformasjon (European Commission 2019). Det er derfor stor usikkerhet i hvordan utviklingen vil foregå fremover.

Vi anbefaler at Forsvaret setter seg i stand til å ta i bruk teknikker innen nevrosymbolisk AI for automatisert prosessering og analyse av data på et tidlig tidspunkt. Dette innebærer først og fremst å ha tilgang til kompetanse på den nyeste utviklingen innen feltet, men siden det er usikkert hvordan disse teknologiene vil bli fulgt opp av akademien og sivil industri bør Forsvaret initiere nødvendig utvikling dersom det viser seg at den sivile utviklingen ikke dekker det militære behovet.

¹¹Technology Readiness Level (TRL) 3, det vil si *proof-of-concept*-nivå.

5 Gjenbruk av modeller

Maskinl ring er avhengig av trening av modeller basert p  relevante data.¹² Trening av modeller er normalt den desidert mest ressurskrevende prosessen i forbindelse med bruk av maskinl ring. I tillegg kan treningen av modeller kreve mye manuelt og kunnskapskrevende menneskelig arbeid, og de n dvendige ressursene for   f  trent en modell kan v re utilgjengelige innenfor Forsvarets rammer.

Det er derfor sv rt nyttig   kunne gjenbruke modeller, b de for   kunne utnytte teknologi som Forsvaret ikke har kapasitet til   bygge internt og for at modeller skal komme til nytte over lengst mulig tid og i s  mange relevante sammenhenger som mulig.

5.1 Eksisterende modeller i ny kontekst

En form for gjenbruk, er   benytte en modell som var trent i  n kontekst og s  anvende den i en annen ulik kontekst. I noen situasjoner er dette trivielt, dersom det kan demonstreres eller p  andre m ter bevises at den nye konteksten er tilstrekkelig lik den gamle, slik at modellen fortsatt er relevant. I andre sammenhenger kan det v re n dvendig med ulike former for tilpasninger for   kunne utnytte den gamle modellen i en ny kontekst.

En form for slik tilpasning, som ble pr vd ut i prosjektet, er overf ringsl ring (*transfer learning*), som tar en eksisterende modell og kombinerer den med menneskelig kunnskap om hvordan den nye konteksten er annerledes enn den opprinnelige, for s    lage en ny modell som er aktuell for den nye konteksten. Et hypotetisk eksempel kan v re at man har tilstrekkelige data til   trene modeller for   klassifisere sivile, men ikke milit re flytyper. Da kan en kombinasjon av modellene fra de sivile flytypene kombineres med milit re data for   bygge modeller som er egnet for de milit re flytypene.

5.2 N r datadistribusjonen endrer seg over tid

En annen form for gjenbruk, er n r en modell skal brukes gjentatte ganger eller kontinuerlig over tid. Dersom distribusjonen¹³ av dataene ikke endrer seg, er dette trivielt fordi den opprinnelige modellen kan benyttes som om ingen tid hadde g tt. Den vil v re like relevant rett etter trening, som etter en dag eller et  r.

Hvis datadistribusjonen derimot endrer seg over tid, kan det v re n dvendig med ulike typer tilpasninger. Hvilke tilpasninger som er n dvendige, er avhengig av en rekke faktorer, spesielt av hvilken type maskinl ringsmodell som er benyttet og ogs  m ten dataene endrer seg. Data kan for

¹²Maskinl ring er metoder som lager modeller basert p  en viss mengde data og s  benytter disse modellene til   si noe om framtidige data.

¹³Distribusjonen av en datamengde beskriver hvilke (kombinasjoner av) verdier som forekommer i dataene og hvor ofte hver verdi forekommer. Maskinl ring baserer seg p  data man har sett p  hittil (treningsdata), og at generalisert kunnskap om treningsdataene vil v re gjeldende for framtidige data (testdata). Dersom distribusjonen av testdataene er annerledes enn for treningsdataene, s  kan egenskapene som l  til grunn for maskinl ringen ikke lenger v re gjeldende.

eksempel endre seg gradvis, med jevne eller ujevne mellomrom eller bare i helt spesielle situasjoner.

Et klassisk eksempel på at data endrer seg over tid, er systemer for å kanalisere relevante data til brukere. Dette er hyppig brukt i det sivile, for eksempel i reklame, nyhetsoppdateringer og nettbutikker. I en slik sammenheng, kan grunnlaget for hvem som bør se hvilken informasjon endre seg med ulike tider i døgnet/uka/måneden/året, som følge av en spesiell nyhetshendelse eller andre viktige omstendigheter.

Slike brukerspesifikke informasjonsoppdateringer kan også være relevante i militær sammenheng, for eksempel for å levere kritisk informasjon til beslutningstakere for en best mulig situasjonsforståelse. Da kan for eksempel øvelser, nye økonomiske rammer og endringer i forholdet til andre nasjoner føre til periodiske eller permanente endringer i grunnlaget for valg av riktig informasjon som skal kanaliseres til hvilke brukere.

Hvor raskt datadistribusjonen endrer seg og måten den endrer seg på, er avgjørende faktorer som påvirker både hvor ofte resultatene må reevalueres, for å sjekke om modellen trenger å oppdateres, og hvor hyppige og omfattende tilpasninger av modellen som er nødvendige for å opprettholde en tilstrekkelig ytelse.

Online læring er metoder som trener maskinlæringsmodeller over tid (Opland 2020), i den hensikt å utnytte akkumulert kunnskap samtidig som modellen justeres tilstrekkelig til å ta høyde for ny kunnskap fra de siste dataene som kommer inn. Det finnes mange hensyn å ta for å få dette riktig, og metoder for å ivareta disse hensynene.

5.3 Konklusjoner og anbefalinger

For å kunne utnytte automatiseringspotensialet som ligger i maskinlæring, er det essensielt at modellene som ligger til grunn er av tilstrekkelig god kvalitet. Trening av slike modeller fra bunnen viser seg imidlertid å kreve såpass store ressurser at det er usikkert om Forsvaret selv vil være i stand til å gjøre dette. Det vil derfor være gunstig for evnen til å utnytte maskinlæring å kunne gjenbruke modeller.

Som vi har påpekt tidligere i rapporten, vil ønskelige egenskaper ved en gitt stordataløsning være sterkt avhengig av spesifikasjonen av den aktuelle stordataløsningen og hvilke krav som stilles til den. På samme måte, vil ulike anvendelser av maskinlæring påvirke både valg av maskinlæringsmetoder og dermed også hva slags data som må samles inn, og hvor stor innsats som trengs for å trene nyttige maskinlæringsmodeller basert på disse dataene. Selv med oppbygging av kompetanse på gjenbruk av modeller, vil det derfor uansett kreves en spesifikk innsats for å få utnyttet maskinlæring for en spesifikk anvendelse.

Vi anbefaler at Forsvaret sikrer seg tilgang på oppdatert kompetanse på gjenbruk av modeller, for å utnytte, justere eller erstatte eksisterende modeller ved behov, i møte med data som endrer seg over tid og for å tilpasse eksisterende modeller til Forsvarets spesifikke behov.

6 Konklusjon

Forsvaret må forvente at antallet stordatap problemer øker i takt med den forventede økende tilgangen på data, det være seg nye data fra noen av Forsvarets tallrike strømmende sensorer, allerede innsamlede data som kan inneholde informasjon som er essensiell i en beslutningssituasjon eller data åpent tilgjengelig på Internett. Derfor bør Forsvaret så raskt som mulig sette seg selv i en posisjon der de kan sørge for at de nødvendige verktøyene for å håndtere disse datamengdene kan etableres. Uansett om Forsvaret selv skal designe slike løsninger eller dette gjøres av samarbeidspartnere, bør Forsvaret være i stand til å spesifisere hvilke egenskaper løsninger bør ha for å svare til de forskjellige stordataproblerne som dukker opp, og også bedømme hvorvidt en foreslått løsning i tilstrekkelig grad svarer til spesifikasjonene. Det er god tilgjengelighet på kostnadsfrie komponenter av høy kvalitet som er klare til å kunne brukes i stordataløsninger, slik at det ligger teknologisk til rette for at slike løsninger kan etableres relativt raskt.

Også for kunnskapsgrafer er det god tilgang på modne og robuste verktøy, både gjennom åpen kildekode og som kommersielle produkter. Det har potensiale til å kunne gi kraftige og fleksible verktøy for å legge til rette for analyse av data sammenstilt fra forskjellige kilder.

Det finnes et utall teknikker som inngår i automatisert prosessering og analyse av data, og maskinlæring, inkludert dyplæring, er en spesielt populær teknikk som har hatt stor utvikling de siste tiårene. Disse teknikkene har imidlertid en svakhet ved at det er vanskelig for en beslutningstaker å forstå modellene som ligger til grunn, noe som kan gjøre det vanskelig å stole på resultatene. Når det skal tas beslutninger som potensielt handler om liv og død, er mangelen på forklarbarhet problematisk. Nevrosymbolske tilnærminger er en måte å møte denne utfordringen på ved å kombinere maskinlæring med teknikker som innebærer logisk resonnering og eksplisitt modellering av kunnskap på en måte en beslutningstaker kan forstå.

Maskinlæring er avhengig av å trene modeller basert på relevante data. Slik trening er normalt den desidert mest ressurskrevende prosessen i forbindelse med bruk av maskinlæring, noe som gjør det svært nyttig å kunne gjenbruke modeller. På den måten kan ressursene brukes på en effektiv måte og Forsvaret kan ta disse teknikkene i bruk selv i tilfeller der de ikke har ressurser til å trene veldig komplekse modeller fra bunnen av.

Vi mener at tiden er moden for Forsvaret å ta stordatateknologier og -teknikker i bruk, og at Forsvaret må sørge for å ha tilstrekkelig tilgang på kompetanse for å kunne gjennomføre automatisert prosessering og analyse av data slik at man kan nærme seg målet om informasjonsoverlegenhet. Vi anbefaler derfor at:

- Forsvaret setter opp en stordataløsning for et utvalgt stordataprobblem for å få erfaring med hvordan slike problemer bør beskrives, hvordan løsninger bør spesifiseres og hvordan løsninger kan realiseres. På den måten vil Forsvaret også få verdifull innsikt i hvilken kompetanse som er nødvendig for å gjennomføre en slik prosess. For å sikre god forankring til problemet som skal løses bør arbeidet ledes av miljøet i Forsvaret som eier den aktuelle problemstillingen, men det følger ikke nødvendigvis av dette at det er Forsvaret selv om også skal gjennomføre oppsett av den tekniske løsningen.
- Forsvaret tester bruk av kunnskapsgrafer i et utvalgt analysemiljø for å se om et slikt verktøy gir verdi i form av bedre og raskere analyseresultater.

-
-
- Forsvaret setter seg i stand til å ta i bruk teknikker innen nevrosymbolisk AI for automatisert prosessering og analyse av data på et tidlig tidspunkt. Dette innebærer først og fremst å ha tilgang til kompetanse på den nyeste utviklingen innen feltet, men siden det er usikkert hvordan disse teknologiene vil bli fulgt opp av akademia og sivil industri bør Forsvaret initiere nødvendig utvikling dersom det viser seg at den sivile utviklingen ikke dekker det militære behovet.
 - Forsvaret sikrer seg tilgang på oppdatert kompetanse på gjenbruk av modeller til maskinlæring slik at slike teknikker kan tas raskt i bruk så snart behovet er der.

Dette er tiltak som reduserer risikoen relatert til å anskaffe stordataløsninger, og etter vårt syn vil Forsvaret ved å gjennomføre disse ta et viktig steg i retning av å kunne løse sine nåværende og framtidige stordatapoblemer.

Referanser

- Brandes, U., Freeman, L. C. & Wagner, D. (2014), *Handbook of graph drawing visualization*, CRC Press, chapter Social Networks, pp. 805–839.
- Bulao, J. (2021), 'How much data is created every day in 2021?'. Besøkt 25.10.2021.
URL: <https://techjury.net/blog/how-much-data-is-created-every-day/>
- Cyganiak, R., Wood, D. & Lanthaler, M. (2014), RDF 1.1 concepts and abstract syntax, W3C recommendation, W3C.
URL: <https://www.w3.org/TR/rdf11-concepts/>
- De Nooy, W., Mrvar, A. & Batagelj, V. (2018), *Exploratory Social Network Analysis with Pajek: Revised and Expanded Edition for Updated Software*, Structural Analysis in the Social Sciences, 3 edn, Cambridge University Press.
- Direktoratet for forvaltning og IKT (2019), 'Veileder for tilgjengeliggjøring av åpne data'.
URL: <https://doc.difi.no/data/veileder-apne-data/>
- European Commission (2019), 'Ethics guidelines for trustworthy AI'. Independent High Level Expert Group on Artificial Intelligence set up by the European Commission.
- Forsvarsdepartementet (2016), 'Kampkraft og bærekraft. Langtidsplan for forsvarssektoren'. Stortingsproposisjon 151S (2015-2016).
- Forsvarsdepartementet (2018), 'Strategi for forskning og utvikling for forsvarssektoren'.
- Forsvarsdepartementet (2020), 'Evne til forsvar – vilje til beredskap. Langtidsplan for forsvarssektoren'. Stortingsproposisjon 14S (2020-2021).
- Garcez, A. d. & Lamb, L. C. (2020), 'Neurosymbolic AI: the 3rd Wave', *arXiv preprint arXiv:2012.05876*.
- Gilbert, S. & Lynch, N. (2002), 'Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services', *Acm Sigact News* **33**(2), 51–59.
- Gunning, D. & Aha, D. (2019), 'DARPA's explainable artificial intelligence (XAI) program', *AI Magazine* **40**(2), 44–58.
- Halvorsen, J. & Hansen, B. J. (2020), Exploring data reuse using a big data infrastructure, FFI-rapport 20/02840, Forsvarets forskningsinstitutt.
- Hansen, B. J. (2021), Utilizing Apache MiNiFi to stream data from tactical sensors, FFI-Notat 21/01565, Forsvarets forskningsinstitutt.
- Heath, T. & Bizer, C. (2011), *Linked data: Evolving the web into a global data space*, Morgan & Claypool Publishers.
- Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutiérrez, C., Kirrane, S., Labra Gayo, J. E., Navigli, R., Neumaier, S., Ngonga Ngomo, A.-C., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J. F., Staab, S. & Zimmermann, A. (2021),

-
-
- Knowledge Graphs*, number 22 in ‘Synthesis Lectures on Data, Semantics, and Knowledge’, Morgan & Claypool.
URL: <https://kgbook.org/>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. et al. (2021), ‘Highly accurate protein structure prediction with AlphaFold’, *Nature* **596**(7873), 583–589.
- Kahneman, D. (2011), *Thinking, Fast and Slow*, Farrar, Straus and Giroux.
- Kautz, H. (2020), ‘Robert S. Englemore Memorial Lecture: The Third AI Summer’.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012), ‘Imagenet classification with deep convolutional neural networks’, *Advances in neural information processing systems* **25**, 1097–1105.
- Laney, D. (2001), 3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical report, META Group.
- Lex Fridman, D. K. (2020), ‘Daniel Kahneman: Thinking Fast and Slow, Deep Learning, and AI’.
URL: <https://lexfridman.com/daniel-kahneman/>
- Lister, M. (2021), ‘31 mind-boggling instagram stats & facts for 2022’. Besøkt 05.11.2021.
URL: <https://www.wordstream.com/blog/ws/2017/04/20/instagram-statistics>
- Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G., Horowitz, M. C., Isbell, C., Kitano, H., Levy, K., Lyons, T., Mitchell, M., Shah, J., Sloman, S., Vallor, S. & Walsh, T. (2021), ‘Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report’.
- Meier, A. & Kaufmann, M. (2019), NoSQL databases, in ‘SQL & NoSQL databases’, Springer, pp. 201–218.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013), Distributed representations of words and phrases and their compositionality, in ‘Advances in neural information processing systems’, pp. 3111–3119.
- NATO Ministers of Defence (2021), ‘Summary of the NATO Artificial Intelligence Strategy’. Meeting of NATO Ministers of Defence - Brussels, 21-22 October 2021.
- Opland, E. J. (2020), Machine learning in a streaming context, FFI eksternnotat 20/00679, Forsvarets forskningsinstitutt.
- Preece, A. D., Braines, D., Cerutti, F., Furby, J., Hiley, L., Kaplan, L., Law, M., Russo, A., Srivastava, M., Vilamala, M. R. et al. (2021), Coalition situational understanding via explainable neuro-symbolic reasoning and learning, in ‘Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III’, Vol. 11746, International Society for Optics and Photonics.
- Reding, D. & Eaton, J. (2020), ‘Science & technology trends 2020-2040’, *NATO Science & Technology Organization, Office of the Chief Scientist, Brussels, Belgium* .

-
-
- Regjeringen Solberg (2020), 'Nasjonal strategi for kunstig intelligens'. Kommunal- og moderniseringsdepartementet.
- Reinsel, D., Gantz, J. & Rydning, J. (2017), Data Age 2025: The Evolution of Data to Life-Critical. Don't Focus on Big Data; Focus on the Data That's Big, Technical report, International Data Corporation (IDC).
- Senger, H. & Geyer, C. (2016), 'Parallel and distributed computing for big data applications', *Concurrency and Computation: Practice and Experience* **8**(28), 2412–2415.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T. et al. (2018), 'A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play', *Science* **362**(6419), 1140–1144.
- Stolpe, A., Hansen, B. J. & Halvorsen, J. (2016), Semantikkbasert informasjonsforvaltning for Forsvaret-sluttrapportering av aktiviteten på semantiske teknologier i FFI-prosjekt Informasjons-og integrasjonstjenester i INI, FFI-rapport 17/00012, Forsvarets forskningsinstitutt.
- Stolpe, A., Hansen, B. J. & Halvorsen, J. (2019), Stordatasystemer og deres egenskaper, FFI-rapport 18/01676, Forsvarets forskningsinstitutt.
- Stolpe, A., Hansen, B. J., Halvorsen, J. & Opland, E. J. (2020), Experimenting with a big data infrastructure for multimodal stream processing, FFI-rapport 20/00480, Forsvarets forskningsinstitutt.
- U.S. Army DEVCOM Army Research Laboratory Public Affairs (2020), 'AI research helps soldiers navigate complex situations'. Besøkt 05.11.2021.
URL: https://www.army.mil/article/240917/ai_research_helps_soldiers_navigate_complex_situations
- Wojcicki, S. (2020), 'Youtube at 15: My personal journey and the road ahead'. Besøkt 25.10.2021.
URL: <https://blog.youtube/news-and-events/youtube-at-15-my-personal-journey/>
- Xing, T., Vilamala, M. R., Garcia, L., Cerutti, F., Kaplan, L., Preece, A. & Srivastava, M. (2019), DeepCEP: Deep complex event processing using distributed multimodal information, in '2019 IEEE International Conference on Smart Computing (SMARTCOMP)', IEEE, pp. 87–92.

Om FFI

Forsvarets forskningsinstitutt ble etablert 11. april 1946. Instituttet er organisert som et forvaltningsorgan, med særskilte fullmakter underlagt Forsvarsdepartementet.

FFIs formål

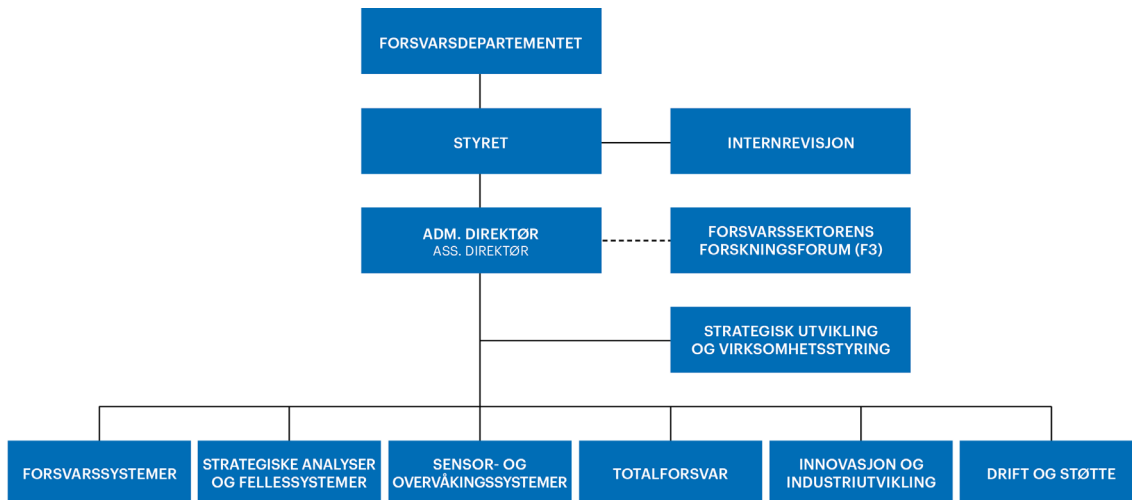
Forsvarets forskningsinstitutt er Forsvarets sentrale forskningsinstitusjon og har som formål å drive forskning og utvikling for Forsvarets behov. Videre er FFI rådgiver overfor Forsvarets strategiske ledelse. Spesielt skal instituttet følge opp trekk ved vitenskapelig og militærteknisk utvikling som kan påvirke forutsetningene for sikkerhetspolitikken eller forsvarsplanleggingen.

FFIs visjon

FFI gjør kunnskap og ideer til et effektivt forsvar.

FFIs verdier

Skapende, drivende, vidsynt og ansvarlig.



Forsvarets forskningsinstitutt
Postboks 25
2027 Kjeller

Besøksadresse:
Instituttveien 20
2007 Kjeller

Telefon: 63 80 70 00
Telefaks: 63 80 71 15
Epost: post@ffi.no

Norwegian Defence Research Establishment (FFI)
P.O. Box 25
NO-2027 Kjeller

Office address:
Instituttveien 20
N-2007 Kjeller

Telephone: +47 63 80 70 00
Telefax: +47 63 80 71 15
Email: post@ffi.no