

# Nevrosymbolisk AI og Agentisk AI

## Momentliste S01–S11

ELMED219 / BMED365

Universitetet i Bergen

Våren 2026

- 1 Grunnleggende paradigmer
- 2 Kunnskapsgrafer og ontologier
- 3 Case: Nevrosymbolisk gliomdiagnose
- 4 Agentisk AI

# S01: Kontrastere symbolsk og konneksjonistisk AI

## Symbolsk AI (GOFAI):

- “Good Old-Fashioned AI”
- Eksplisitte regler og symboler
- Logisk resonnering
- Kunnskapsrepresentasjon
- Ekspert-systemer

## Styrker:

- Forklarbar
- Presise resonnementer
- Kodifiserer ekspertkunnskap

## Konneksjonistisk AI:

- Nevrale nettverk, dyplæring
- Lærer fra data
- Subsymbolisk representasjon
- Mønstergjenkjenning
- “Black box”

## Styrker:

- Lærer komplekse mønstre
- Håndterer støy og usikkerhet
- Skalerer med data

## Historisk perspektiv

AI har svingt mellom disse paradigmene. Nå: Kan vi kombinere dem?

## S02: Forklare konseptet nevrosymbolisk integrasjon

Nevrosymbolisk AI = det beste fra to verdener

Hovedidé:

- Kombiner **nevralt** (læring fra data, mønstergjenkjenning)
- med **symbolisk** (resonnering, kunnskap, forklarbarhet)

Integrasjonsmønstre:

- ① **Neural → Symbolic:** Nevralt nettverk produserer symboler for resonnering
- ② **Symbolic → Neural:** Symbolisk kunnskap guider nevralt nettverk
- ③ **Hybrid:** Tett integrasjon der begge informerer hverandre

Eksempel: Medisinsk diagnose

**Neval:** CNN analyserer MR-bilde, identifiserer tumor-features

**Symbolisk:** Regelbasert klassifikasjon iht. WHO-kriterier

**Integritt:** Kombinerer bildefunn med kliniske regler for diagnose

## S03: Beskrive hva en kunnskapsgraf er

### Kunnskapsgraf (Knowledge Graph):

- Strukturert representasjon av kunnskap som en **graf**
- Noder = **entiteter** (personer, sykdommer, legemidler)
- Kanter = **relasjoner** ("behandles med", "forårsaker")

### Trippel-representasjon:

(Subjekt, Predikat, Objekt)  
*(Diabetes, behaves\\_med, Metformin)*

### Anvendelser:

- Google Knowledge Graph
- Medisinsk beslutningsstøtte
- Legemiddelinteraksjoner

### Fordeler:

- Maskinlesbar kunnskap
- Resonnering over relasjoner
- Koblet data (Linked Data)

### Teknologier

RDF, OWL, SPARQL – standarder for kunnskapsrepresentasjon og spørring

## S04: Kjenne til medisinske ontologier

**Ontologi = formell spesifikasjon av begreper og relasjoner i et domene**

**Viktige medisinske ontologier/terminologier:**

Navn	Beskrivelse	Anvendelse
SNOMED CT	Kliniske termer, 350 000+ begreper	EPJ, klassifisering
ICD-10/11	Internasjonal sykdomsklassifikasjon	Rapportering, statistikk
LOINC	Laboratorieundersøkelser	Labresultater
RxNorm	Legemidler	Forskrivning
WHO CNS 2021	CNS-tumorklassifikasjon	Gliomdiagnose

### WHO 2021 CNS-klassifikasjon

Integrerer histologiske og molekylære kriterier. Eksempel: IDH-mutant astrocytom krever både histologi **og** IDH-mutasjonsstatus.

## S05: Diskutere fordeler med nevrosymbolisk AI i medisin

### Potensielle fordeler:

- ① **Forklarbarhet:** Symbolisk komponent gir tolkbare resonnementer
- ② **Dataholdighet:** Symbolisk kunnskap kompenserer for lite data
- ③ **Robusthet:** Regelbaserte begrensninger forhindrer absurde prediksjoner
- ④ **Oppdaterbarhet:** Ny medisinsk kunnskap legges til som regler/fakta
- ⑤ **Samsvar med retningslinjer:** Kode WHO-kriterier, guidelines direkte
- ⑥ **Validering:** Sjekk om prediksjoner er konsistente med kjent kunnskap

### Hvorfor dette er viktig i medisin

Medisin er et domene der ekspertkunnskap er rik, forklarbarhet er kritisk, og feil kan være fatale. Ren dyplæring er ofte utilstrekkelig.

# S06: Nevrosymbolisk AI for gliomdiagnostikk

## Case study fra Lab 3, Notebook 08:

### Problemet:

- Gliom-klassifikasjon krever både **bildeanalyse** (MRI) og **molekylære markører** (IDH, MGMT, 1p/19q)
- WHO 2021 krever integrasjon av flere informasjonskilder

### Nevrosymbolisk løsning:

- ① **Neural komponent:** CNN segmenterer tumor i MRI (BraTS)
- ② **Symbolisk komponent:** WHO-kriterier kodet som kunnskapsgraf
- ③ **Integrasjon:** Regelbasert klassifikator bruker CNN-output + molekylære markører

### Resultat

Diagnose som er både **bildebasert** (CNN) og **kriteriekonform** (WHO) – med forklarbar resonnering.

## S07: Forklare hvordan kunnskapsgrafer kan validere nevrale prediksjoner

### Validering av CNN-prediksjoner med symbolsk kunnskap:

#### Scenario:

- CNN predikerer: "Glioblastom med IDH-mutasjon"
- Kunnskapsgraf inneholder: "Glioblastom er per definisjon IDH-villtype (WHO 2021)"
- Konflikt detektert!

#### Valideringsmekanismer:

- ① Konsistenssjekk: Er prediksjonen logisk konsistent?
- ② Regelvalidering: Oppfyller prediksjonen nødvendige kriterier?
- ③ Plausibilitetssjekk: Er kombinasjonen av funn realistisk?

#### Verdi

Fanger opp potensielle CNN-feil **før** de når kliniker. Øker tilliten til systemet.

## S08: Definere agentisk AI og dens kjerneegenskaper

**Agentisk AI = AI som handler autonomt mot mål**

**Kjerneegenskaper:**

- ① **Autonomi:** Handler uten kontinuerlig menneskelig input
- ② **Målrettet:** Jobber mot definerte mål
- ③ **Planlegging:** Bryter ned komplekse mål i deloppgaver
- ④ **Verktøybruk:** Kan bruke eksterne verktøy (API, databaser)
- ⑤ **Refleksjon:** Evaluerer egen fremgang, justerer strategi
- ⑥ **Minne:** Husker kontekst over flere interaksjoner

**LLM-baserte agenter:**

- LLM som "hjerne" – resonnerer og planlegger
- Verktøy-kall (function calling) for handlinger
- Eksempler: AutoGPT, LangChain agents, Claude tool use

## S09: Beskrive hvordan en AI-agent kan orkestrere klinisk arbeidsflyt

### Eksempel: Agentisk AI for gliomutredning

Scenario: Lege ber om “Utred ny MR for mulig gliom”

Agenten kan:

- ① **Hent data:** Koble til PACS, laste ned MR-bilder
- ② **Analyser:** Kjør BraTS-segmentering, beregn tumorvolum
- ③ **Litteratur:** Søk PubMed for relevante studier
- ④ **Biobank:** Sjekk om molekylær analyse er tilgjengelig
- ⑤ **Klassifiser:** Bruk WHO-kriterier for tentativ diagnose
- ⑥ **Rapporter:** Generer strukturert rapport for MDT-møte

### Orkestrering

Agenten **koordinerer** flere AI-systemer og datakilder mot ett klinisk mål – uten at legen må gjøre hvert steg manuelt.

## S10: Forklare konseptet human-in-the-loop i agentiske systemer

### HITL i agentisk AI:

- Selv autonome agenter trenger **menneskelig tilsyn**
- Spesielt viktig for beslutninger med store konsekvenser

### Implementeringsmønstre:

- ❶ **Godkjenningspunkter:** Agent pauser for godkjenning før kritiske handlinger
- ❷ **Konfidensterskel:** Menneske involveres når agent er usikker
- ❸ **Audit trail:** Alle handlinger logges for review
- ❹ **Overriding:** Menneske kan avbryte eller korrigere agent

### I medisinsk kontekst

Agent kan samle data og foreslå diagnose, men **legen** tar endelig beslutning. Agenten er en "superkraftig assistent", ikke en autonom beslutningstaker.

# S11: Diskutere etiske utfordringer med autonome AI-agenter i helsevesenet

## Etiske utfordringer:

- ① **Ansvar:** Hvem er ansvarlig når en agent gjør feil?
  - Utvikler? Sykehus? Lege som "slapp løs" agenten?
- ② **Autonomi og kontroll:** Hvor mye autonomi er for mye?
  - Risiko for utilsiktet handling
- ③ **Transparens:** Kan vi forstå hvorfor agenten tok beslutningene?
- ④ **Sikkerhet:** Agenter med tilgang til systemer = angrepsflate
- ⑤ **Dehumanisering:** Risiko for å redusere menneskelig kontakt i helse
- ⑥ **Overtillit:** "Automation bias" – blindt støle på agenten

## Prinsipper for ansvarlig agentisk AI

Gradvis autonomi, robust HITL, klar ansvarsfordeling, omfattende testing, kontinuerlig overvåking

# Oppsummering: Nevrosymbolisk AI og Agentisk AI

## Nevrosymbolisk AI:

- S01–S02: Kombinerer nevralgt (læring) og symbolisk (resonnering)
- S03–S04: Kunnskapsgrafer og medisinske ontologier
- S05–S07: Fordeler i medisin, case gliomdiagnose, validering

## Agentisk AI:

- S08: Autonome, målrettede AI-systemer med planlegging
- S09: Orkestrering av kliniske arbeidsflyter
- S10–S11: HITL og etiske utfordringer

## Lab 3, Notebook 08

Utforsk en detaljert case study som demonstrerer nevrosymbolisk gliomklassifikasjon og agentisk orkestrering.