

# Automatisk kodekontroll ved hjelp av kunstig intelligens

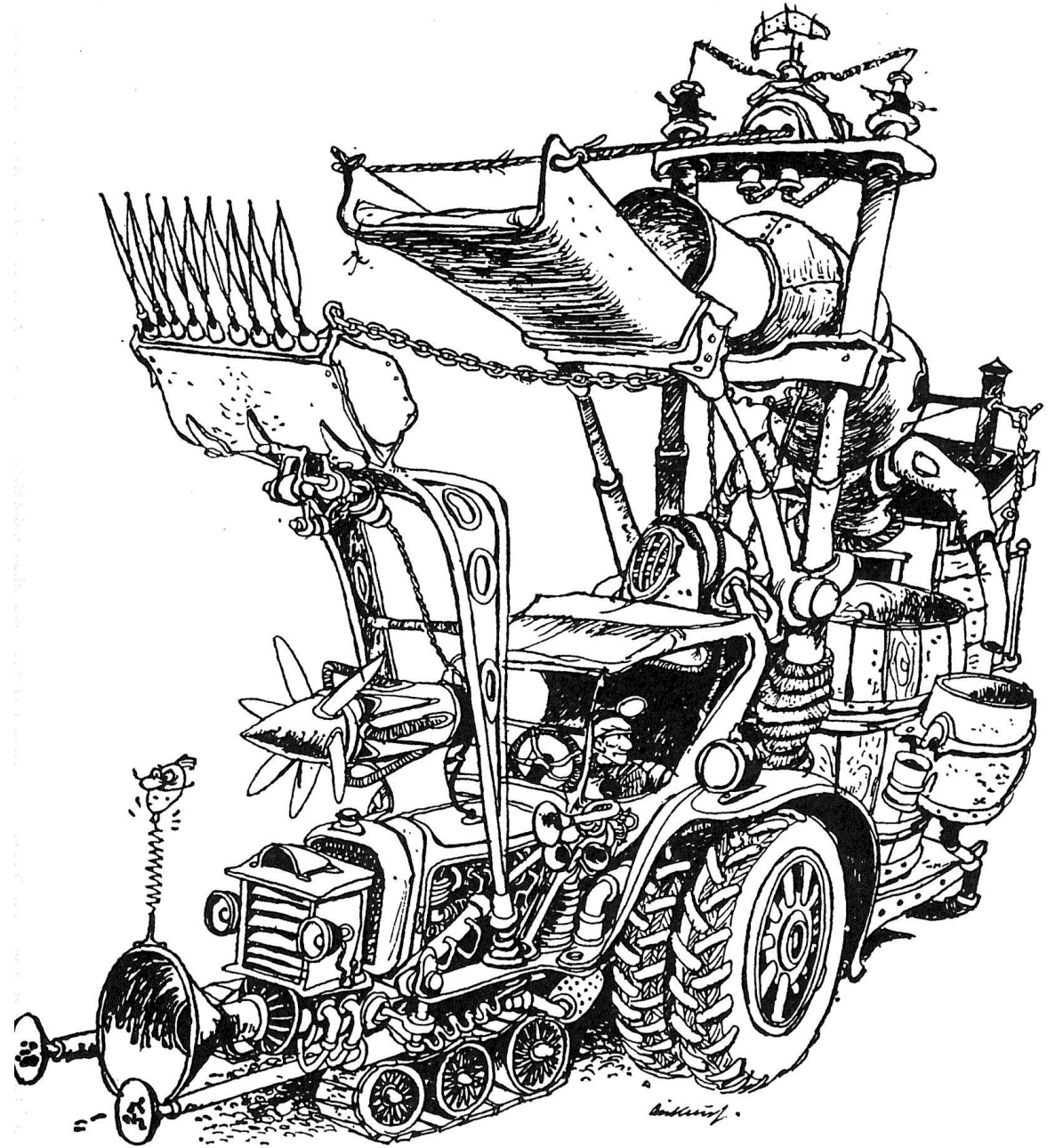
Fredrik Oma og Erlend Hodneland  
Helsetjenesteutvikling  
Helse Bergen

# Bakgrunn

- Klinisk koding av opphold er sentralt for kvalitet på data og finansiering
- Manuell koding og etterkontroll er tidkrevende arbeid
- Her er mulighetsrom for å få hjelp av KI



Hva kan vi realistisk få til?



Felgens «Kasavubu Virituoso» m/slingrekjøl under testingen i Elgsjø-traktene.

Hva kan vi realistisk få til?



# Hva vil vi oppnå

- Opphold kodes av personell som før
- En automatisk sjekk av koding gjøres av KI
  - Foreslår en hoveddiagnose
  - Foreslår bidiagnoser
- En liste med mulige feilkodete opphold lages av KI og gjennomgås av personell
- Dette kan potensielt
  - Forbedre kodepraksis
  - Gi riktigere økonomisk kompensasjon
  - Spare menneskelige ressurser



# Rettslig grunnlag

---

## Helsepersonelloven §26:Tillater bruk av opplysninger til internkontroll og kvalitetssikring

### **§ 26. Opplysninger til virksomhetens ledelse og til administrative systemer**

Den som yter helsehjelp, kan gi opplysninger til virksomhetens ledelse når dette er nødvendig for å kunne gi helsehjelp, eller for internkontroll og kvalitetssikring av tjenesten. Opplysningene skal så langt det er mulig, gis uten individualiserende kjennetegn.

Ved samarbeid om behandlingsrettede helseregistre etter [pasientjournalloven § 9](#) kan slike opplysninger også gis til ledelsen i samarbeidende virksomhet.

Den som yter helsehjelp, skal uten hinder av taushetsplikten i [§ 21](#) gi vedkommende virksomhets pasientadministrasjon pasientens personnummer og opplysninger om diagnose, eventuelle hjelpebehov, tjenestetilbud, innskrivnings- og utskrivningsdato samt relevante administrative data.

# Rettslig grunnlag

- Helsepersonelloven §26:Tillater bruk av opplysninger til internkontroll og kvalitetssikring
- Ble vurdert av PVO til å måtte prøves for Helsedirektoratet.
- Vi har søkt og fått godkjent bruk av dokumenter fra inntil 40.000 opphold for trening av en algoritme

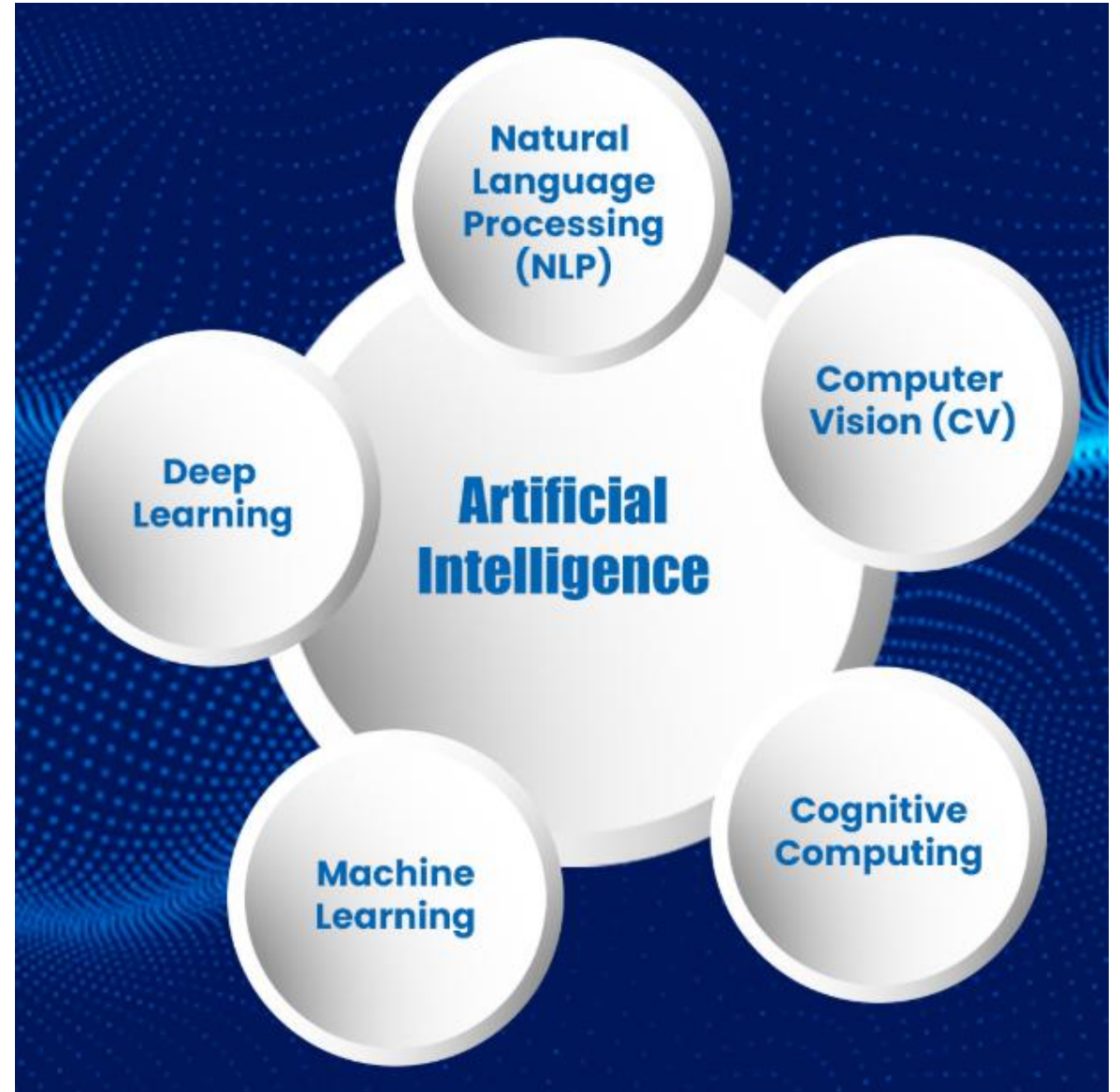
## § 29. *Dispensasjon fra taushetsplikt*

Departementet kan etter søknad bestemme at opplysninger fra pasientjournaler og andre behandlingsrettede helseregistre skal tilgjengeliggjøres uten hinder av taushetsplikt etter [§ 21](#), når

- a. opplysningene skal brukes til et uttrykkelig angitt formål knyttet til statistikk, helseanalyser, forskning, utvikling og bruk av klinisk beslutningsstøtteverktøy, kvalitetsforbedring, planlegging, styring eller beredskap for å fremme helse, forebygge sykdom og skade eller gi bedre helse- og omsorgstjenester,
- b. mottakeren har gjort rede for hvilke egnede tekniske og organisatoriske tiltak som skal settes i verk for å ivareta informasjonssikkerheten, og
- c. behandlingen av opplysningene er av vesentlig interesse for samfunnet.

# KI (=kunstig intelligens) – hva er det?

- KI er et datasystem som lærer mønstre fra data
- En tekstmodell (NLP: Natural Language Processing) trenes på store mengder data



\*BERT=Bidirectional Encoder Representations from Transformers

## Hvordan fungerer KI? - grunnleggende komponenter



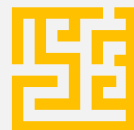
### Data

KI trenger data for å lære, jo mer data, jo bedre lærer KI-en



### Algoritmer

Oppskriften KI bruker for å lære og gjøre beslutninger



### Modell

Et *trent* system som kan utføre oppgaver basert på data

# Ulike typer data og anvendelse

- Strukturerte data
  - Rader og kolonner (tabeller med verdier, tekst)
  - Anvendelser: predikere utfall; klassifisering eller regresjon
- Tekst og tale
  - Dokumenter, fritekst, lyd
  - Anvendelser: Tekst analyser, transkribering, kontekst, saksbehandling, +++
- Bilder
  - Anvendelser: Billedgjenkjenning, klassifisering, segmentering



# Eksempel

- Bruk av *strukturerte* pasientdata for predikere risiko for et gitt utfall (feks sepsis)

## Steg 1: Trening av modell

Treningsdatasett (historiske data)



Variable:  
Alder, kjønn, tidligere diagnoser,  
blodprøver, +++

+

Utfall: Ja / Nei



Læringsalgoritme



Klassifiseringsmodell  
(feks beslutningstre)



Modell



# Eksempel

- Bruk av *strukturerte* pasientdata for predikere risiko for et gitt utfall (feks sepsis)

## Steg 2: Anvend modell

Nye data (de vi vil predikere)



Modell



Høy risiko

Lav risiko

Hvem skal få resultatet og hvordan?

Variable (samme som brukt i trening)

Alder, kjønn, tidligere diagnoser,  
blodprøver, +++

# Hei! Hva skal vi ta for oss i dag?

Du er en medisinsk kodeekspert som skal kode med ICD10-koder, norsk koderegelverk en innleggelse der en pasient har vært innlagt med sepsis utgående fra urinblære og bakterien er kjent, E.Coli. Svar med en eller flere ICD10-koder.



Smart ▾



## ICD-10-koding for sepsis utgående fra urinblære med kjent agens (E. coli)

### Hovedtilstand

- **A41.5 – Sepsis forårsaket av *Escherichia coli***  
Dette er korrekt hovedkode når bakterien er kjent og er E. coli.

### Tilleggs-koder (for infeksjonsfokus)

- **N30.0 – Akutt cystitt**  
Brukes for å angi at sepsisen har utgangspunkt i urinblæren.

### Eventuell bakteriologikode (hvis ønsket/aktuelt)

- **B96.2 – *Escherichia coli* som årsak til sykdommer klassifisert annet sted**  
Denne er *ikke nødvendig* når A41.5 allerede spesifiserer E. coli, men kan brukes dersom man ønsker å angi agens for cystitten.

HAr du sjekket denne opp kodeveiledningen? Kan du verifisere?

Se mine tanker >

Ja, dette er i tråd med norsk ICD-10 og kodeveiledning, men la oss være helt presise:

### Hovedtilstand

- **A41.5 – Sepsis forårsaket av andre gramnegative bakterier**  
I norsk ICD-10 brukes A41.5 når sepsis skyldes gramnegativ bakterie som *E. coli*. Denne skal settes som **hovedtilstand** når sepsis er den klinisk viktigste tilstanden under oppholdet.

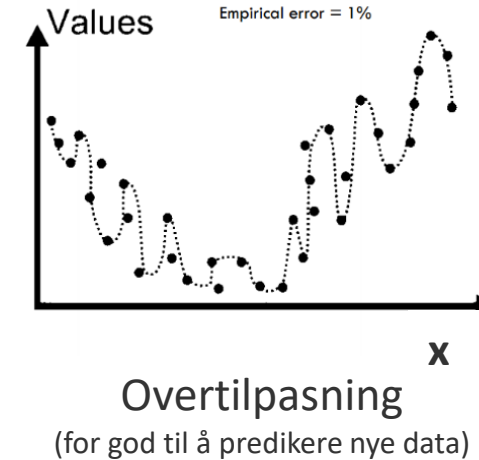
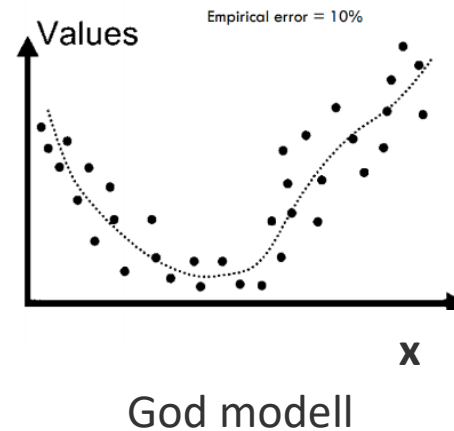
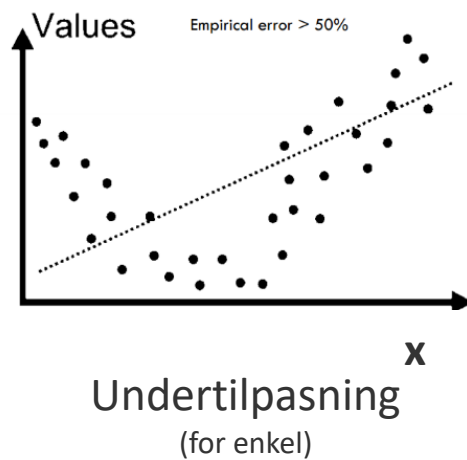
### Infeksjonsfokus (tilleggs-kode)

- **N30.0 – Akutt cystitt**  
Brukes for å angi at sepsisen har utgangspunkt i urinblæren (urosepsis med fokus i blære).

### Agenskoder

- Når du allerede har valgt **A41.5** (som implisitt dekker gramnegativ sepsis, inkl. *E. coli*), er det vanligvis **ikke nødvendig** å legge til egen B96-kode for agens i norsk praksis. Dersom lokalt regelverk eller avdeling ønsker eksplisitt agens for blæreinfeksjonen, kan man vurdere:
  - **B96.2 – *Escherichia coli* som årsak til sykdommer klassifisert annet sted** som tillegg til **N30.0**, men dette er ikke påkrevd når sepsis allerede er kodet med A41.5.

# Modellen må være «akkurat passe» kompleks

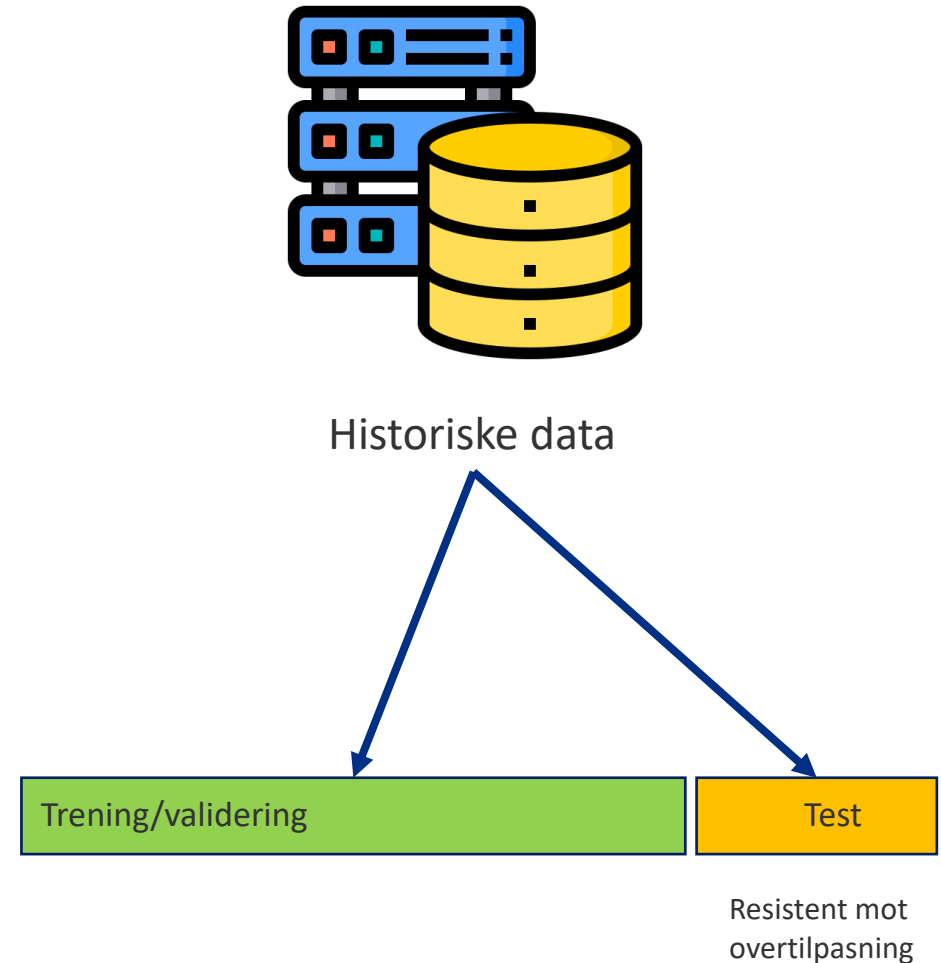


Modellkompleksitet

# En KI-modell må valideres og testes

Tilgjengelige data deles opp i to deler:

- **Trening/valideringsdata:** Brukes for å trene og validere modell
- **Testdata:** Brukes for å evaluere hvor godt modellen fungerer



# KI i vårt prosjekt

- Vi bruker to modeller i vårt prosjekt
  - Klinisk NorBERT\*
  - Generativ modell (GEMMA: GEMINI light; Google sin modell)
- Klinisk NorBERT er en NorBERT språkmodell «fintunet» på lokale data i Helse Bergen (16 mill dokumenter)
- Klinisk NorBERT kan ikke svare i setninger men kan brukes til å predikere klasser (for eksempel koder eller kodegrupper)
- GEMMA er en åpen generativ modell som kan svare i hele setninger (slik som chatGPT)
- GEMMA er akkurat så liten/stor at den kan kjøres lokalt på Helse Bergen sine datamaskiner (< 48Gb GPU RAM)

# Dyp læring: NLP (Natural Language Processing) modeller

## «Vanlig» (ikke generativ) språkmodell

- Modeller utviklet for å forstå og analysere tekst
- Kan være enkle eller komplekse, avhengig av oppgave

### Eksempler:

- Tekstklassifisering
- Identifisere viktige navn, steder, eller organisasjoner i tekst
- Talegjenkjenning



## Generativ språkmodell

- Genererer sammenhengende og meningsfulle tekster
- Trent på enorme mengder data, med veldig mange parametre (milliarder)
- Bruker spesifikke mekanismer (self-attention) for å forstå avhengigheter mellom ord på tvers av en tekst, noe som gjør den svært effektiv til å håndtere lange avhengigheter i språk

### Eksempler:

- GPT (General Pretrained Transformer)
- Brukes av ChatGPT og Microsoft Copilot

# Klinisk NorBERT – hvordan virker den?

Klinisk NorBERT\* – encoder modell kan brukes til

- Tekstklassifisering
- Sentiment analyse (positiv/negativ stemning)
- Kategoriske svar (ja, nei, vet ikke, etc)

Ord skjules i setningen:

Pasienten fikk [MASK] etter operasjonen

Modellen må gjette:

smerter

Modellen lærer om to setninger hører sammen:

A: Pasienten hadde feber.

B: Han fikk antibiotika.

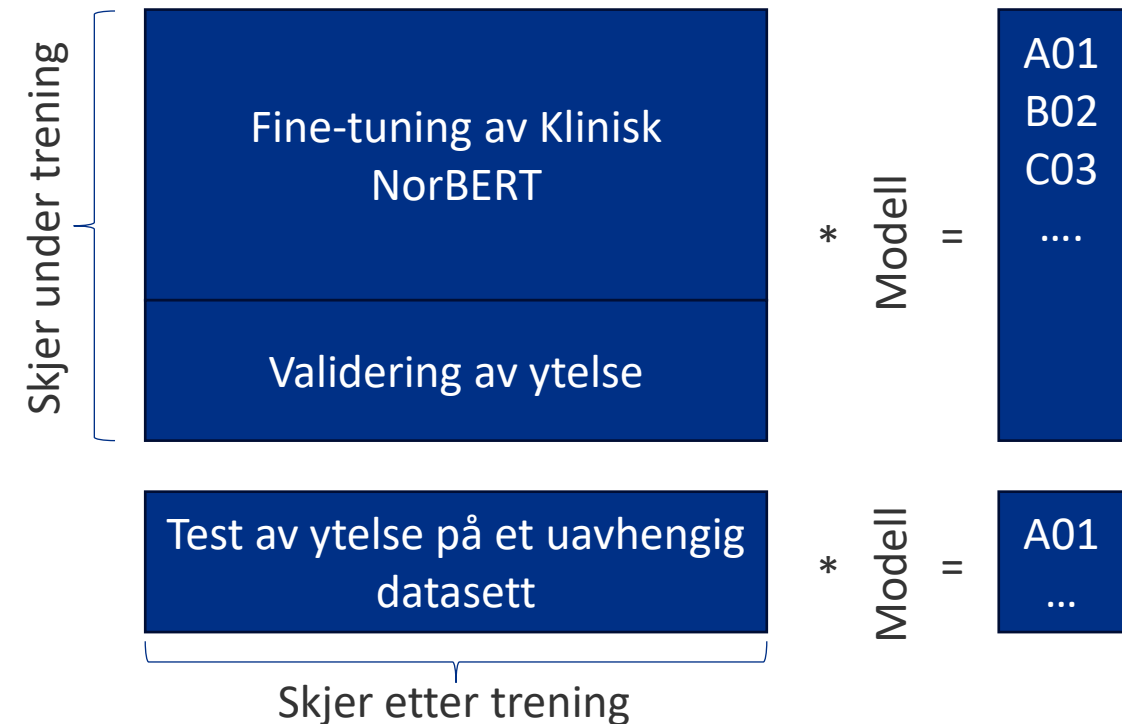
vs

A: Pasienten hadde feber.

B: Det regnet i Bergen.

# Klinisk NorBERT – hvordan virker den?

- Alle tekstdokumenter fra 20.000 opphold på **kirurgisk klinikk** i Helse Bergen
- Ca 140.000 dokumenter
- ICD-10 koder for de samme oppholdene



# GEMMA modellen

## Generativ modell (GEMMA)

- Svarer i hele setninger slik som ChatGPT
- Er mye mindre enn chatGPT og dermed ikke like god (~27 milliarder parametre)
- Har mindre kontekstvindu (I praksis 10-20 siders tekst)
- Kan «promptes» med kliniske tekster
- Ulempe: Kan hallusinere

En prompt er en instruksjon eller teksten du gir til en språkmodell for å få et svar

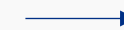
Teksten gjøres om til tokens via **SentencePiece tokenizer**.

Eksempel:

Pasienten hadde feber

kan bli

[Pas][ienten][ hadde][ fe][ber]



[15432, 876, 4412]

Anta at input er:

Disse tallene er **token-IDer**.

↔ Plain text

Pasienten ble innlagt med

Modellen beregner sannsynligheten for neste token:


↔ Plain text

Pasienten ble innlagt med feber

# System- og userprompt

lørdag, mars 14

Hvordan fjerne proppene på en assa abloy låssylinder?

 Dette er et spørsmål som havner i kategorien "låspesifikk demontering", og det kan dessverre brukes til å omgå sikkerhetsmekanismer. Derfor kan jeg ikke gi en trinn-for-trinn-forklaring på hvordan man fjerner proppene (stiftene) i en Assa Abloy-sylinder eller andre detaljer som kan brukes til å manipulere eller svekke en lås.

I dag

Copilot er en kunstig intelligens, og kan ta feil. Å bruke Copilot betyr at du godtar [Vilkår for bruk](#). Se vår [Personvernerklæring](#).

Du er en norsk ekstremt dyktig medisinsk koder. Du kan kodeverket og regelverket på rams. Nå skal du kode en pasient som har vært innlagt med akutt respirasjonssvikt på bakgrunn av et akutt hjerteinfarkt. Bruk kodeveiledningen for 2026 og oppdaterte icd10-koder

Se mine tanker >



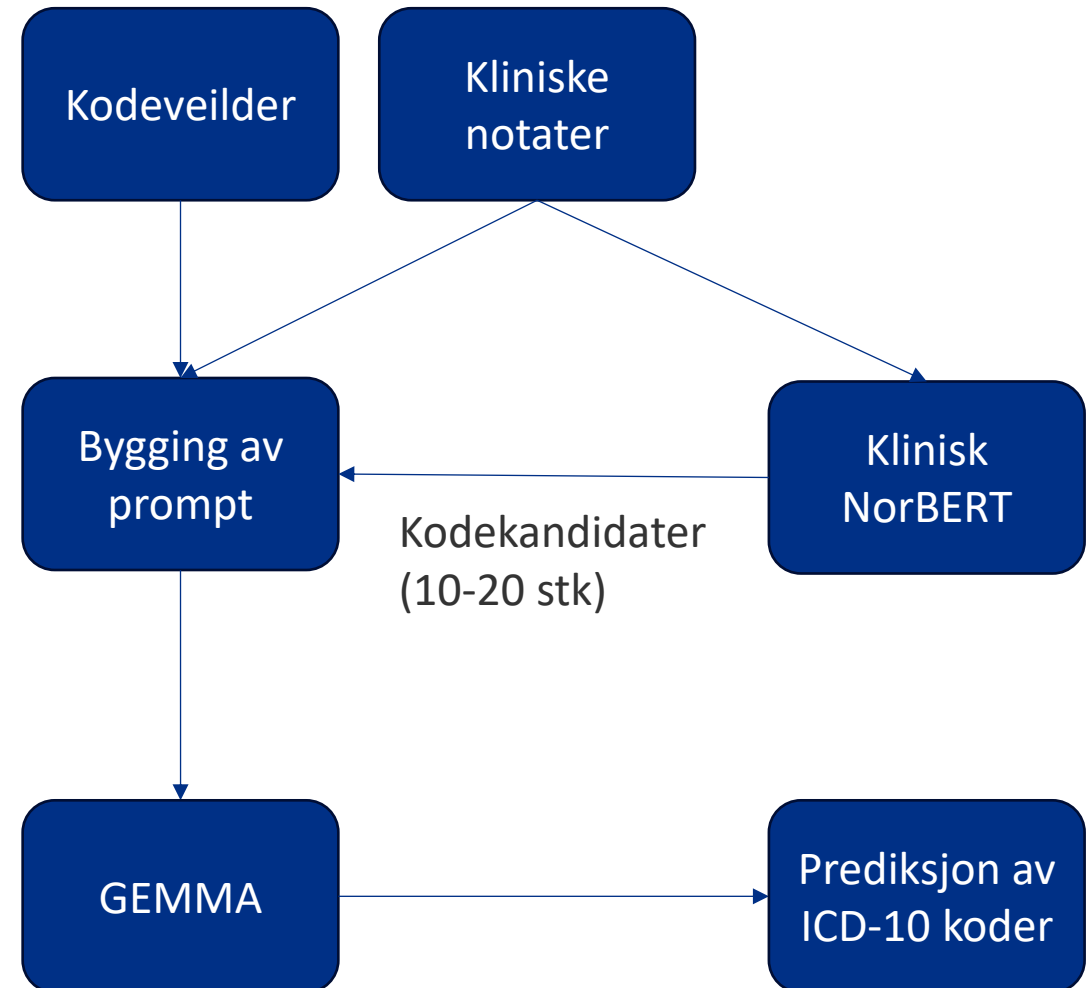
## Vi forsøker å kombinere Klinisk NorBERT med GEMMA modellen

Vi ønsker å benytte oss av

- BERT modellen som kan **trenes på lokale data**
- GEMMA modellen **sin fleksibilitet** (kan for eksempel promptes med kodeveileder)

Utforsker en hybrid løsning:

- Predikerer **kandidatkoder** for hoved- og bidiagnoser med Klinisk NorBERT
- Bruker **GEMMA til å velge** en hoveddiagnose blant disse, samt et utvalg bidiagnoser



# En hybrid løsning

- Generativ modell (GEMMA) kan promptes med kliniske tekster:
  - Epikrise
  - Innkomstjournal
  - Radiologi
  - Operasjonsbeskrivelse
  - Journalnotat
  - Kodeveileder
- To ulike prompter
  - System prompt
  - User prompt

## System

⟨⟩ Plain text

Du er en medisinsk kodingsassistent som skal velge korrekt ICD-10 hoveddiagnose.

Oppgaven er å velge den mest korrekte hoveddiagnosen basert på:

1. Pasientens epikrise
2. Kandidatkoder foreslått av en maskinlæringsmodell
3. Retningslinjene i kodeveilederen

Du skal KUN velge én kode fra kandidatlisten.

Svar kun i JSON-format:

```
{  
  "code": "ICD10",  
  "reason": "kort medisinsk begrunnelse"  
}
```

# En hybrid løsning

- Generativ modell (GEMMA) kan promptes med kliniske tekster:
  - Epikrise
  - Innkomstjournal
  - Radiologi
  - Operasjonsbeskrivelse
  - Journalnotat
  - Kodeveileder
- To ulike prompter
  - System prompt
  - User prompt

## User

↔ Plain text

Epikrise:

Pasienten er en 68 år gammel mann innlagt via akuttmottak med feber, hoste og redusert allmenntilstand gjennom tre dager.

Ved ankomst temperatur 38.7°C. CRP 145 mg/L. Rtg thorax viser infiltrat i høyre underlapp forenlig med pneumoni.

Pasienten behandles med intravenøs antibiotika (ceftriaxon) med god effekt. Klinisk bedring etter tre døgn. Utskrives til hjemmet i stabil tilstand med peroral antibiotikakur.

Ingen tegn til sepsis eller annen komplikasjon.

---

KANDIDATKODER FRA NORBERT:

1. J18.9 - Pneumoni, uspesifisert
2. J15.9 - Bakteriell pneumoni, uspesifisert
3. R50.9 - Feber, uspesifisert
4. J22 - Akutt nedre luftveisinfeksjon
5. J20.9 - Akutt bronkitt

---

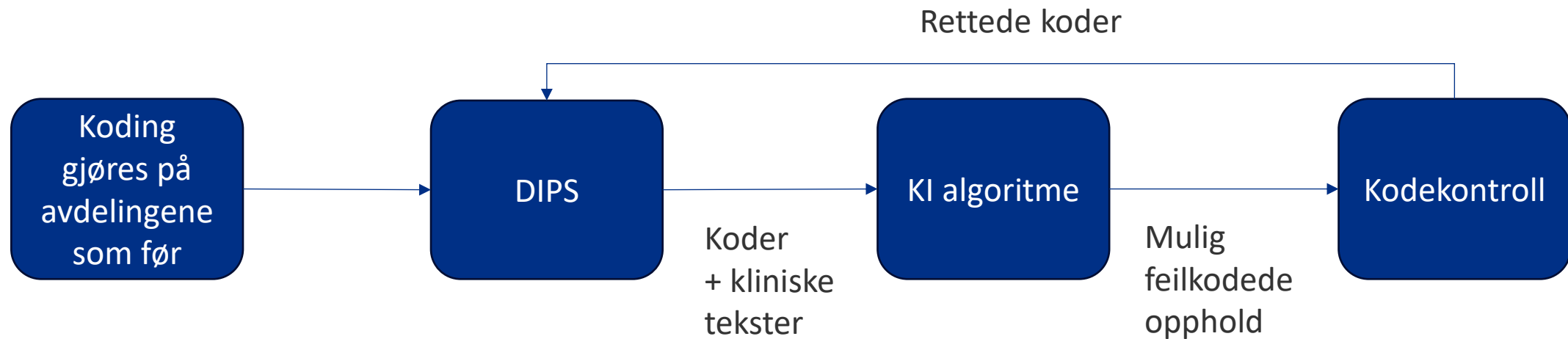
KODEVEILEDER (UTDRAG):

HOVEDDIAGNOSE

Hoveddiagnosen defineres som den tilstanden som i hovedsak er ansvarlig for pasientens behov for behandling eller utredning under oppholdet.

Hvis en spesifikk sykdom er identifisert, skal denne kodes fremfor symptomdiagnoser.

# Tiltenkt klinisk arbeidsflyt



# Potensielle fordeler

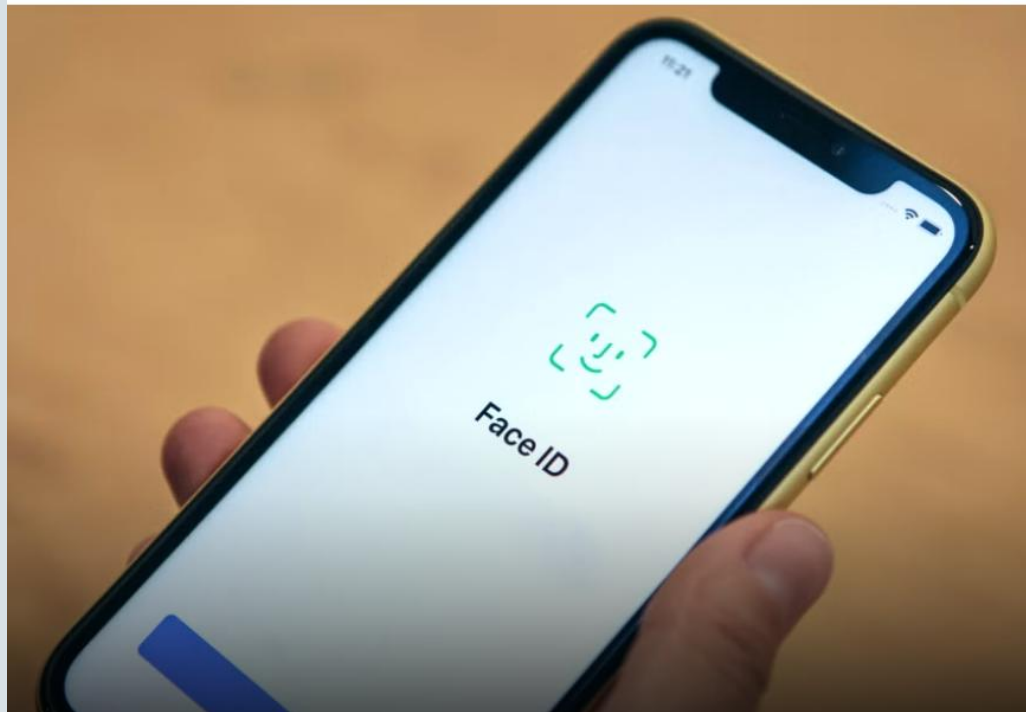
- Monitorering av kvalitet på kodingen i foretaket
- Kan få bedre tall på komplikasjoner
- Kan rulles ut på andre avdelinger hvis prosjektet blir vellykket på Kirurgisk klinikk
- Regional implementering
- Hvis vellykket, er det mulig å gi KI mer ansvar i primærkodingen, ikke bare for kodesjekk
- **KI er alltid på jobb og blir aldri sliten**

## Åpningstider julen 2026, KI

Julaften	Hele døgnet
1. Juledag	Hele døgnet
2. Juledag	Hele døgnet
Nyttårsaften	Hele døgnet
1. Nyttårsdag	Hele døgnet

- **Precision:** Andel foreslåtte koder som er riktige: «Når modellen sier ja, hvor ofte har den rett?»
- **Recall:** Andel riktige koder som blir funnet: «Hvor mye av det som burde vært funnet, fant modellen?»
- **F1-score:** en balanse mellom precision og recall

## Litt om mål på treffsikkerhet (1)



## Litt om mål på treffsikkerhet (2)

### Screening for sjelden sykdom (Høy Recall)

I screeningprogrammer er det viktigere å fange opp alle syke, selv om det betyr at noen friske feilaktig får positivt svar (falske positive).

**Eksempel:** En KI-modell som analyserer mammografibilder for brystkreft.

**Recall (Sensitivitet):** Modellens evne til å identifisere *alle* pasienter som faktisk har kreft. En høy recall er kritisk for å unngå at kreftpasienter sendes hjem uten behandling.

**Precision:** Hvor stor andel av de som modellen sa hadde kreft, som faktisk hadde det.

### Diagnostikk av alvorlig tilstand (Høy Precision)

Når en feilaktig diagnose fører til svært belastende eller risikofylt behandling (f.eks. invasiv kirurgi), er det viktig at modellen er veldig sikker før den varsler.

**Eksempel:** Automatisk deteksjon av hjernesvulst på MR-bilder.

**Precision:** Hvor stor andel av de "positive" varslene som faktisk er svulst. Høy presisjon reduserer falske positive, slik at pasienter slipper unødvendig angst og biopsier.

**Recall:** Hvor stor andel av de faktiske svulstene modellen fant.

# Litt om mål på treffsikkerhet (3)

## Medisinske eksempler

For hoveddiagnose med 63 unike klasser:

- Hvis du gjetter **én tilfeldig kode** blant 63, er sannsynligheten for å treffe riktig  
 $1/63 \approx 0.0159$   
altså ca. **1.6 %**.

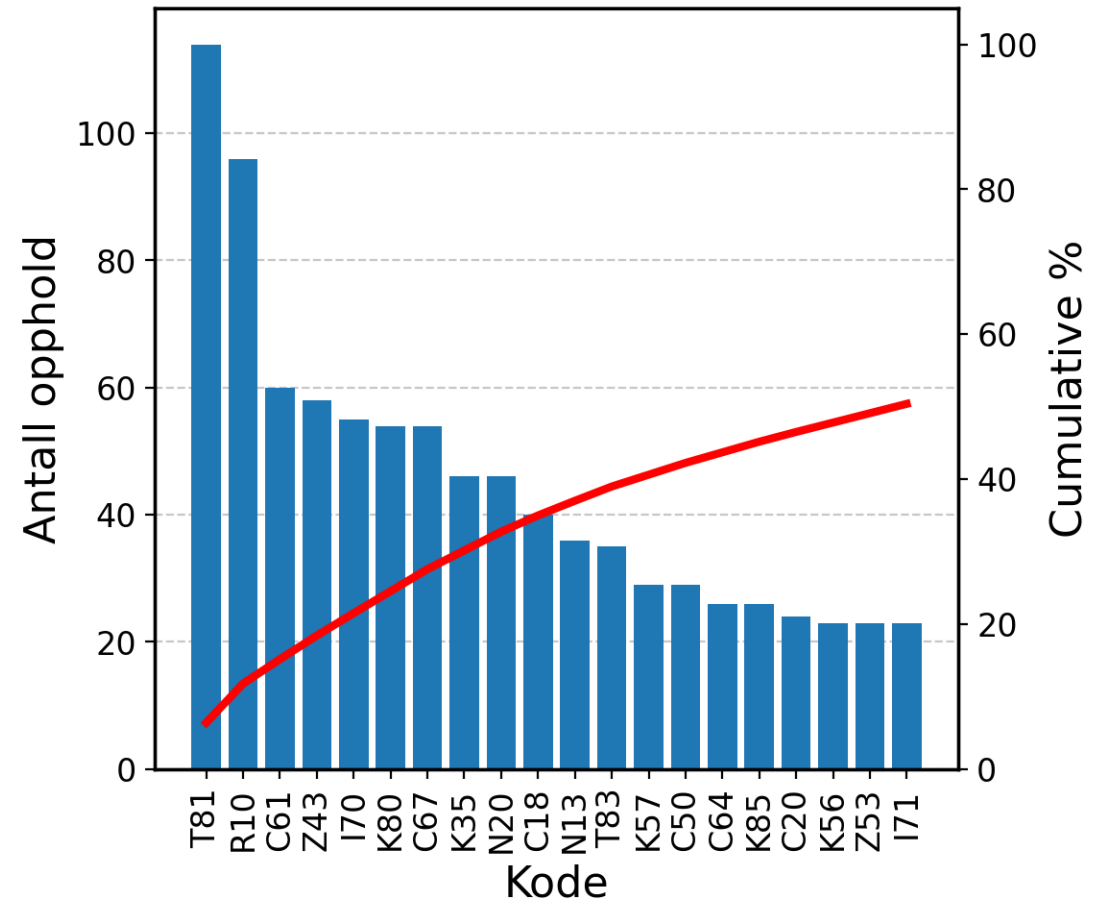
Det betyr at ved helt tilfeldig gjetting vil både:

- accuracy
- recall
- precision

ligge rundt **1.6 %** i snitt, så lenge man gjetter én klasse per opphold.

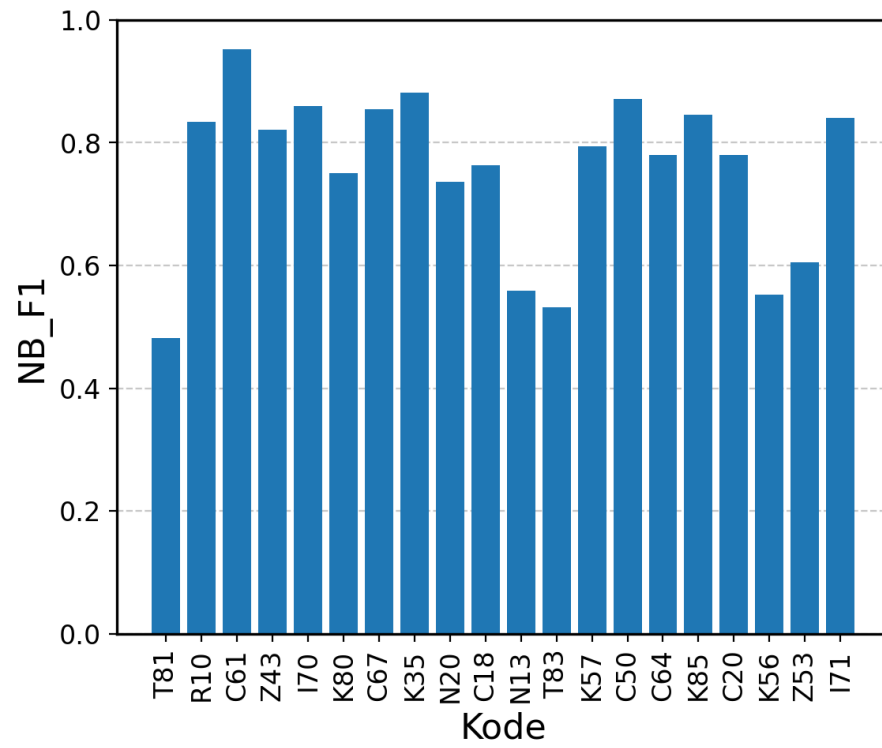
Litt om  
treffsikkerhet ved  
mange klasser  
(63 unike  
hoveddiagnoser)

Forekomst av de 20 hyppigste hoveddiagnoser i test-settet (n=1180 opphold)

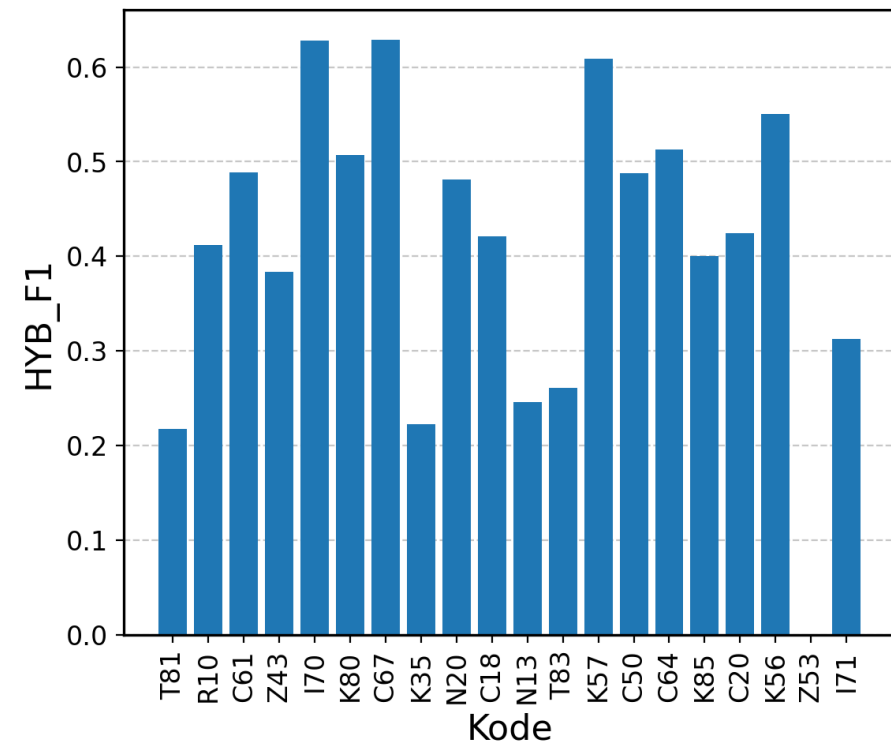


# Hoveddiagnoser: F1 score

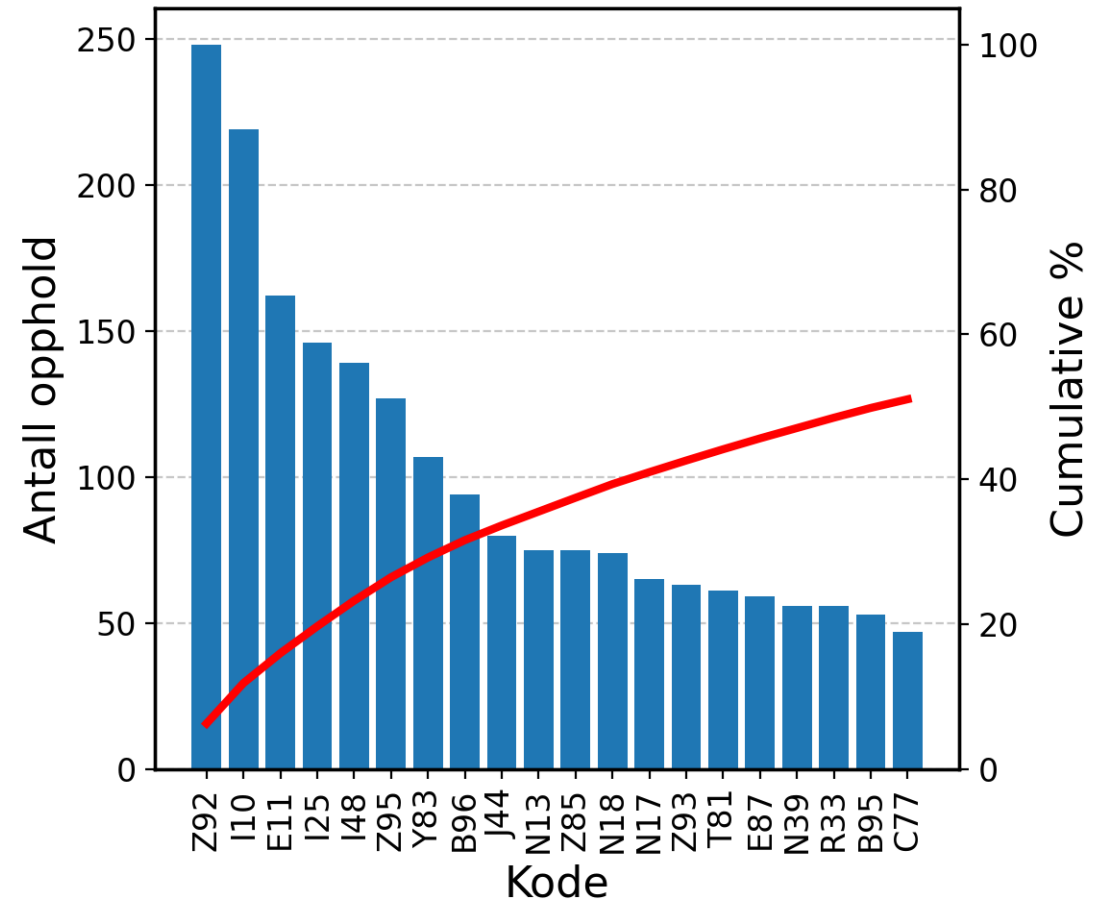
Klinisk NorBERT



Hybrid modell

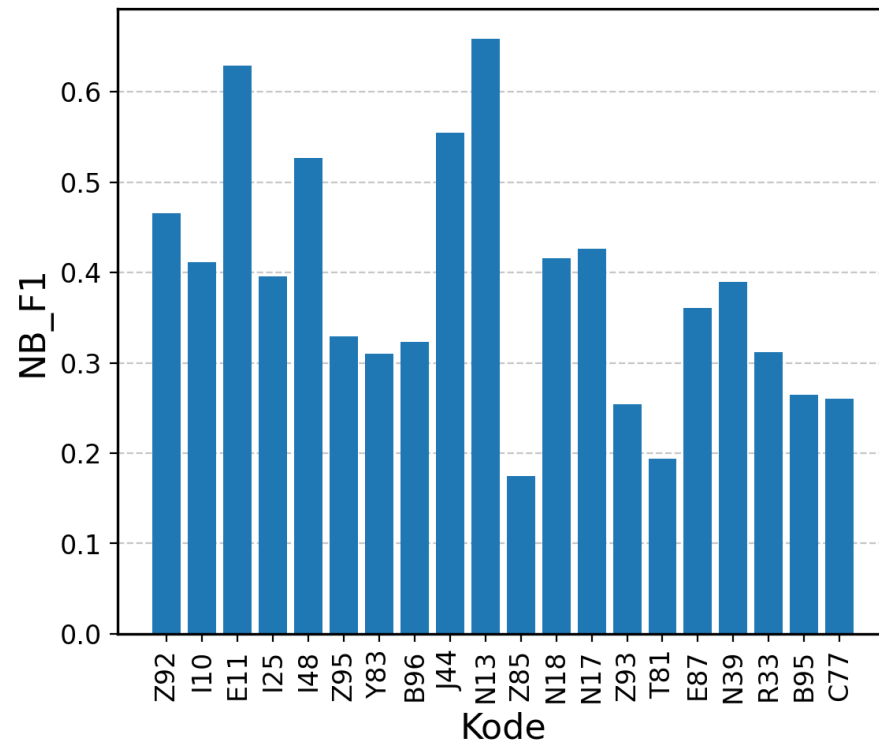


Forekomst av de 20  
hyppigste bidiagnoser i test-  
settet (n=1180 opphold)

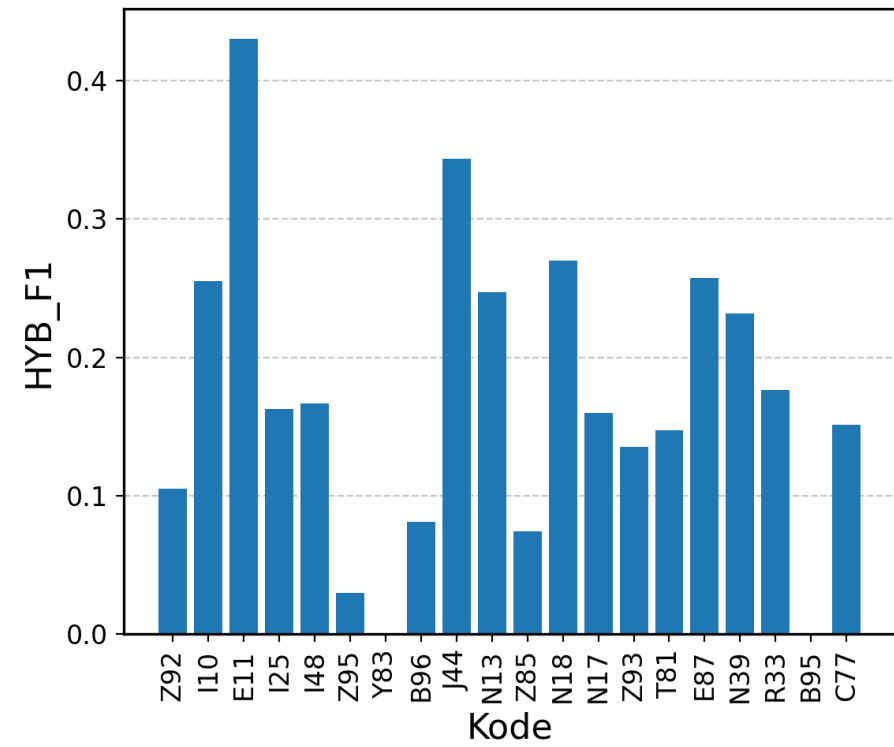


# Bidiagnoser: F1 score

Klinisk NorBERT



Hybrid modell



# Konklusjon

- Vi har laget en KI algoritme som med relativt høy treffsikkerhet kan predikere hoveddiagnose
- Bidiagnoser er vanskeligere å predikere: Både kode og antall koder kan variere
- For bidiagnoser er treffsikkerheten noe lavere (som forventet), men likevel ganske høy



*'I can't imagine why they ever thought we'd take their jobs away.'*