

# AI-etikk og Regulering

## ELMED219: Momentliste A01–A09

ELMED219

Vår 2026

# Oversikt

1 Bioetikk og AI

2 EU AI Act

3 Ansvar og rettferdighet

# A01: Diskutere de fire bioetiske prinsippene i AI-kontekst

## Beauchamp & Childress' fire prinsipper:

### 1. Autonomi (Respekt for selvbestemmelse)

- Informert samtykke til AI-bruk
- Rett til å vite om AI er involvert
- Mulighet til å velge bort AI-beslutninger

### 2. Velgjørenhet (Gjøre godt)

- AI skal forbedre pasientutfall
- Validert og effektiv
- Tilgjengelig for alle

### 3. Ikke-skade (Primum non nocere)

- Unngå skade fra feil prediksjoner
- Minimere bias og diskriminering
- Sikkerhetstesting før klinisk bruk

### 4. Rettferdighet (Rettferdig fordeling)

- Lik tilgang til AI-forbedret helse
- Unngå systematisk diskriminering
- Representativ treningsdata

## Utfordring

Prinsippene kan komme i konflikt – f.eks. effektivitet (velgjørenhet) vs. personvern (autonomi)

## A02: Identifisere typer bias i medisinske AI-systemer

**Bias = systematisk skjevhets som fører til urettferdige utfall**

**Kilder til bias:**

- ① **Historisk bias:** Treningsdata reflekterer historisk urettferdighet
  - Eksempel: Underrepresentasjon av minoriteter i kliniske studier
- ② **Representasjonsbias:** Skjev pasientpopulasjon i treningsdata
  - Eksempel: Modell trent på kun hvite pasienter
- ③ **Målbias:** Feil proxy-variabler
  - Eksempel: Bruk av helsekostnader som mål på sykdom (korrelerer med rase)
- ④ **Evalueringsbias:** Testdata er ikke representativ
- ⑤ **Aggregeringsbias:** En modell for alle, ignorerer subgrupper

**Konsekvens**

Bias kan føre til at AI-systemer gir dårligere helsehjelp til allerede marginaliserte grupper.

# A03: Forklare personvernghensyn ved bruk av LLM

## Personvernutfordringer med LLM:

### Datalekasje til tredjepart:

- Input sendes til LLM-leverandør
- Potensielt brukt til retraining
- Lagres på utenlandske servere

### Memorisering:

- LLM kan ha "lært" persondata
- Kan generere sensitiv info i output

### Identifikasjon:

- Re-identifisering fra tilsynelatende anonyme data
- Prompt kan avsløre kontekst

### Tiltak:

- Aldri del pasientidentifiserbar info med offentlige LLM
- Bruk lokale/sikre løsninger (chat.uib.no)
- Anonymisering og pseudonymisering

## Viktig regel

**Aldri legg inn pasientdata i ChatGPT, Claude eller andre offentlige LLM!**

# A04: Beskrive hovedtrekkene i EU AI Act

## EU AI Act (vedtatt 2024):

- Verdens første omfattende AI-regulering
- Risikobasert tilnærming
- Gjelder alle som tilbyr eller bruker AI i EU

## Hovedprinsipper:

- ① **Risikoklassifisering:** AI-systemer kategoriseres etter risiko
- ② **Krav proporsjonale med risiko:** Høyere risiko = strengere krav
- ③ **Transparens:** Brukere må informeres om AI-bruk
- ④ **Menneskers tilsyn:** Human-in-the-loop for høyrisiko
- ⑤ **Dokumentasjon:** Teknisk dokumentasjon og logging
- ⑥ **CE-merking:** Samsvarsvurdering før lansering

## Tidslinje

Gradvis innføring 2024–2027. Høyrisiko AI-krav gjelder fra 2026.

# A05: Forklare risikoklassifisering i EU AI Act

Fire risikonivåer:

## ① Uakseptabel risiko (FORBUDT)

- Social scoring, manipulasjon, real-time ansiktsgjenkjenning

## ② Høy risiko (STRENGE KRAV)

- Medisinsk utstyr og diagnostikk
- Kritisk infrastruktur, utdanning, arbeidsmarked
- Krever risikovurdering, dokumentasjon, menneskelig tilsyn

## ③ Begrenset risiko (TRANSPARENSKRAV)

- Chatbots, deepfakes
- Må informere om at det er AI

## ④ Minimal risiko (INGEN SPESIFIKKE KRAV)

- Spamfiltre, spill-AI
- Frivillige etiske retningslinjer

Medisinsk AI

De fleste medisinske AI-systemer vil klassifiseres som **høy risiko**.

# A06: Diskutere krav til høyrisiko AI i helsevesenet

## Krav til høyrisiko AI-systemer:

### Før lansering:

- Risikovurdering og -håndtering
- Høy datakvalitet og representativitet
- Teknisk dokumentasjon
- Transparens og brukerveiledning
- Cybersikkerhet
- Samsvarsvurdering (CE-merking)

### Under bruk:

- Automatisk logging
- Menneskelig tilsyn
- Post-market overvåking
- Hendelsesrapportering
- Periodisk revurdering

## Spesifikt for medisinsk AI:

- Koordineres med MDR (Medical Devices Regulation)
- Klinisk evidens og validering kreves
- Krav om forklarbarhet og robusthet

# A07: Reflektere over ansvarsfordeling når AI feiler

## Hvem er ansvarlig når AI tar feil?

### Potensielle ansvarlige:

- **AI-utvikleren:** Design, trening, validering
- **Helseinstitusjonen:** Innkjøp, implementering, opplæring
- **Klinikeren:** Bruker AI, tar endelig beslutning
- **Pasienten:** Samtykker til AI-bruk?

### Problemstillinger:

- AI som “black box” – vanskelig å forstå feil
- Delt ansvar → utvannet ansvar?
- Automatiseringsbias – overtillit til AI

### Rettslig utvikling:

- AI Liability Directive (EU)
- Produktansvar utvidet til AI
- Bevisbyrde kan legges på AI-leverandør

## Hovedprinsipp

Kliniker har fortsatt det endelige ansvaret – AI er et **verktøy**, ikke en erstatning.

# A08: Kjenne til GDPR-relevante aspekter ved AI

## GDPR og AI i helse:

### Nøkkelprinsipper:

- **Lovlig grunnlag:** Samtykke, nødvendighet, berettiget interesse
- **Formålsbegrensning:** Data kun til oppgitt formål
- **Dataminimering:** Kun nødvendige data
- **Lagringsbegrensning:** Ikke lagre lenger enn nødvendig

### Spesielle bestemmelser:

- **Art. 22:** Rett til å ikke bli utsatt for automatiserte beslutninger
- **Art. 9:** Helsedata er “spesiell kategori” – ekstra beskyttet
- **Art. 35:** DPIA (personvernkonsekvensvurdering) for høyrisiko

### Viktig

Pasientdata brukt til AI-trening/bruk krever lovlig grunnlag og ofte DPIA.

# A09: Diskutere algoritmisk rettferdighet (fairness)

## Fairness i ML: Ingen universell definisjon

### Ulike rettferdighetsdefinisjoner:

- ① **Demografisk paritet:** Lik andel positive prediksjoner per gruppe
- ② **Equalised odds:** Lik TPR og FPR per gruppe
- ③ **Calibration:** Konfidensscorer betyr det samme for alle grupper
- ④ **Individual fairness:** Like individer behandles likt

### Problemet:

- Disse definisjonene er ofte **matematisk inkompatible**
- Valg av definisjon er en **verdiladet** beslutning
- "Fair" i én forstand kan være "unfair" i en annen

### Medisinsk eksempel

Skal en risikomodell gi lik score til en 30-åring og 70-åring med samme symptomer? Alder er en risikofaktor – er det diskriminering?

# Oppsummering: AI-etikk og Regulering

## Bioetikk og bias:

- A01: Fire bioetiske prinsipper i AI-kontekst
- A02: Typer bias – historisk, representasjons-, mål-, evaluerings-
- A03: Personvernghensyn ved LLM

## EU AI Act:

- A04–A06: Risikobasert tilnærming, høyrisiko-krav for medisin

## Ansvar og rettferdighet:

- A07: Ansvarsfordeling – komplekst, kliniker har endelig ansvar
- A08: GDPR – spesialbeskyttelse for helsedata
- A09: Algoritmisk rettferdighet – ingen enkel løsning

## Praktisk i Lab 3

Diskuter etiske implikasjoner av LLM, bias-analyse, og personvernghensyn.